Ferdowsi University of Mashhad

Iranian Journal of Numerical Analysis and Optimization

Volume 17, Number 1

March 2027

Serial Number: 40

Ferdowsi University of Mashhad, Iran

In the Name of God

Iranian Journal of Numerical Analysis and Optimization (IJNAO)

This journal is authorized under the registration No. 174/853 dated 1386/2/26 (2007/05/16), by the Ministry of Culture and Islamic Guidance.

Volume 17, Number 1, March 2027

ISSN-Print: 2423-6977, **ISSN-Online:** 2423-6969

Publisher: Faculty of Mathematical Sciences, Ferdowsi University of Mashhad

Published by: Ferdowsi University of Mashhad Press

Printing Method: Electronic

Address: Iranian Journal of Numerical Analysis and Optimization

Faculty of Mathematical Sciences, Ferdowsi University of Mashhad

P.O. Box 1159, Mashhad 91775, Iran.

Tel. : +98-51-38806222 , **Fax:** +98-51-38807358

E-mail: ijnao@um.ac.ir

Website: http://ijnao.um.ac.ir

This journal is indexed by:

- SCOPUS
- ZbMATH Open
- ISC
- DOAJ
- Civilica
- Magiran
- Mendeley
- Academia.edu
- Linkedin

- The Journal granted the International degree by the Iranian Ministry of Science, Research, and Technology.

Iranian Journal of Numerical Analysis and Optimization

Volume 17, Number 1, March 2027

Ferdowsi University of Mashhad - Iran

©2027 All rights reserved. Iranian Journal of Numerical Analysis and Optimization

Iranian Journal of Numerical Analysis and Optimization

Director

M. H. Farahi

Editor-in-Chief

Ali R. Soheili

Managing Editor

M. Gachpazan

EDITORIAL BOARD

Abbasbandi, Saeid*

(Numerical Analysis)

Imam Khomeini International University, Iran.

e-mail: abbasbandy@ikiu.ac.ir

Abdi, Ali*

(Numerical Analysis)

University of Tabriz, Iran.

e-mail: a_abdi@tabrizu.ac.ir

Area, Iván*

(Numerical Analysis)

Universidade de Vigo, Spain.

e-mail: area@uvigo.es

Babaie Kafaki, Saman*

(Optimization)

Semnan University, Iran.

e-mail: sbk@semnan.ac.ir

Babolian, Esmail*

(Numerical Analysis)

Kharazmi University, Iran.

e-mail: babolian@khu.ac.ir

Cardone, Angelamaria*

(Numerical Analysis)

Università degli Studi di Salerno, Italy.

e-mail: ancardone@unisa.it

Dehghan, Mehdi*

(Numerical Analysis)

Amirkabir University of Technology, Iran.

e-mail: mdehghan@aut.ac.ir

Effati, Sohrab*

(Optimal Control & Optimization)

Ferdowsi University of Mashhad, Iran.

e-mail: s-effati@um.ac.ir

Emrouznejad, Ali*

(Operations Research)

Aston University, UK.

e-mail: a.emrouznejad@aston.ac.uk

Farahi, Mohammad Hadi*

(Optimal Control & Optimization)

Ferdowsi University of Mashhad, Iran.

e-mail: farahi@um.ac.ir

Gachpazan, Mortaza**
(Numerical Analysis)
Ferdowsi University of Mashhad, Iran.
e-mail: gachpazan@um.ac.ir

Ghanbari, Reza*
(Operations Research)
Ferdowsi University of Mashhad, Iran.
e-mail: rghanbari@um.ac.ir

Hadizadeh Yazdi, Mahmoud*
(Numerical Analysis)
Khaje-Nassir-Toosi University of
Technology, Iran.
e-mail: hadizadeh@kntu.ac.ir

Hojjati, Gholamreza*
(Numerical Analysis)
University of Tabriz, Iran.
e-mail: ghojjati@tabrizu.ac.ir

Hong, Jialin*
(Scientific Computing)
Chinese Academy of Sciences (CAS), China.
e-mail: hjl@lsec.cc.ac.cn

Karimi, Hamid Reza*
(Control)
Politecnico di Milano, Italy.
e-mail: hamidreza.karimi@polimi.it

Khojasteh Salkuyeh, Davod*
(Numerical Analysis)
University of Guilan, Iran.
e-mail: khojasteh@guilan.ac.ir

Lohmander, Peter*
(Optimization)
Swedish University of Agricultural Sciences, Sweden.
e-mail: Peter@Lohmander.com

Lopez-Ruiz, Ricardo*
(Complexity, nonlinear models)
University of Zaragoza, Spain.
e-mail: rilopez@unizar.es

Mahdavi-Amiri, Nezam*
(Optimization)
Sharif University of Technology, Iran.
e-mail: nezamm@sina.sharif.edu

Mirzaei, Davoud*
(Numerical Analysis)
University of Uppsala, Sweden.
e-mail: davoud.mirzaei@it.uu.se

Omrani, Khaled*
(Numerical Analysis)
University of Tunis El Manar, Tunisia.
khaled.omrani@issatso.rnu.tn

Soheili, Ali Reza*
(Numerical Analysis)
Ferdowsi University of Mashhad, Iran.
e-mail: soheili@um.ac.ir

Soleimani Damaneh, Majid*
(Operations Research and Optimization, Finance, and
Machine Learning)
University of Tehran, Iran.
e-mail: m.soleimani.d@ut.ac.ir

Toutounian, Faezeh*

(Numerical Analysis)

Ferdowsi University of Mashhad, Iran.

e-mail: toutouni@um.ac.ir

Türkyılmazoğlu, Mustafa*

(Applied Mathematics)

Hacettepe University, Turkey.

e-mail: turkyilm@hacettepe.edu.tr

Vahidian Kamyad, Ali*

(Optimal Control & Optimization)

Ferdowsi University of Mashhad, Iran.

e-mail: vahidian@um.ac.ir

Xu, Zeshui*

(Decision Making)

Sichuan University, China.

e-mail: xuzeshui@263.net

Vasagh, Zohreh

(English Text Editor)

Ferdowsi University of Mashhad, Iran.

This journal is published under the auspices of Ferdowsi University of Mashhad

* Full Professor

** Associate Professor

We would like to acknowledge the help of Miss Narjes khatoon Zohorian in the preparation of this issue.

Letter from the Editor–in–Chief

I would like to welcome you to the Iranian Journal of Numerical Analysis and Optimization (IJNAO). This journal has been published two issues per year and supported by the Faculty of Mathematical Sciences at the Ferdowsi University of Mashhad. The faculty of Mathematical Sciences with the centers of excellence and the research centers is well-known in mathematical communities in Iran.

The main aim of the journal is to facilitate discussions and collaborations between specialists in applied mathematics, especially in the fields of numerical analysis and optimization, in the region and worldwide. Our vision is that scholars from different applied mathematical research disciplines pool their insight, knowledge, and efforts by communicating via this international journal. In order to assure the high quality of the journal, each article is reviewed by subject-qualified referees. Our expectations for IJNAO are as high as any well-known applied mathematical journal in the world. We trust that by publishing quality research and creative work, the possibility of more collaborations between researchers would be provided. We invite all applied mathematicians especially in the fields of numerical analysis and optimization to join us by submitting their original work to the Iranian Journal of Numerical Analysis and Optimization.

We would like to inform all readers that the Iranian Journal of Numerical Analysis and Optimization (IJNAO), has changed its publishing frequency from "Semiannual" to a "Quarterly" journal since January 2023. The four journal issues per year will be published in the months of March, June, September, and December. One of our goals is to continue to improve the speed of both the review and publication processes, while try continuing to publish the best available international research in numerical analysis and optimization, with the high scientific and publication standards that the journal is known for.

Ali R. Soheili

Editor-in-Chief

Contents

Adaptive hybrid numerical schemes with singularity treatment for third order Lane–Emden-type equations 1
M. Fazli

Unraveling the relationship between COVID-19 viral load, disease severity, and innate immune response: Insights from mathematical modeling and sensitivity analysis 25
M. Lotfi, A. Jabbari, R. Rahbarghazi and H. Kheiri

A stable and high-order accurate D-RBF-PU method via hybrid kernels 45
M. Esmaeilbeigi and R. Cavoretto

Behavioral modeling and control strategies for Rumor-induced customer dynamics on digital platforms 72
S. Belhdid, H. Hassouni, O. Aarabate and O. Balatif

Understanding the impact of vaccination on Omicron variants 109
G.P. Sahu, A.S. Thakur, C.S. Sikarwar and H. Singh

A modified school bullying model with a recovery compartment: Stability analysis and optimal control 145
A. Sasmito, C. Miftahurrizki K H and R. Maeni

Modeling the transmission dynamics of meningococcal meningitis under hospitalization effects 177
I.O. Joshua, K. Kiruthika, M. Gümüş, A. Raza, E. Chandrakumar and N. Jeeva

A predator-prey model incorporating fear effect, prey vigilance costs, and predator hunting cooperation: Analysis and simulations 217
A. Akinsanya

Algebraic properties of the new polyconvolution operator related to the Laplace, Fourier cosine integral transforms and applications 247
M.K. Nguyen, L.T.P. Bui and V.T. Tran

Mathematical modeling and optimal control of hepatitis B virus dynamics . . 273
L. El Houari, M. Elkaf and J. Danane

Memory-driven fractional dynamics of pneumonia transmission with prophylactic intervention 290
K. Soni, S. Kumar, S. Kumawat and A.K. Sinha

A conjugate gradient algorithm based on a combined descent scheme for two-layer neural network training and unconstrained optimization 332
C. Tebbal, M. Laouar, R. Ziadi and A.E. Mehamdia

Enhancing conjugate gradient performance via a diagonal quasi-Newton strategy: Applications in adaptive filtering and power allocation 359

N. Bidabadi

Mathematical analysis of Mpox transmission using an SAIR-SI model with optimal control strategies 380
T. Janan, F. Fatmawati and A. Hasan

Stability analyses of block hybrid methods for numerical solutions of differential equations 416
O. Adedire and U.W. Sirisena

An approach to find the optimal design of a 3D damping set for a damped wave equation in spherical coordinates 457
H. Alimorad, A. Fakharzadeh Jahromi and M. Goodarzi

Stability and sensitivity analysis of a nonlinear compartmental model for banking risk contagion 494
S. Ibnoukady, A. Meskaf and K. Allali

Mathematical modeling, global stability, sensitivity analysis, and optimal control of merchant adoption dynamics in e-commerce 529
H. Hassouni, A. El Bhih, O. Aarabate and O. Balatif

Global dynamics of a fractional-order Hantavirus model with awareness-driven protective behavior 570
Z. Yaagoub

An appropriate fractional Narayana polynomials neural network method for dental caries infection model 603
H. Hassani, Z. Avazzadeh, A. Turan-Dincel and H. Farahmand

Iranian Journal of Numerical Analysis and Optimization
Vol. 17, No. 1, 2027, pp 1–24
https://doi.org/10.22067/ijnao.2026.97038.1812
https://ijnao.um.ac.ir/

CC BY NC

Research Article

Adaptive hybrid numerical schemes with singularity treatment for third order Lane–Emden-type equations

M. Fazli*,

Abstract

This paper introduces two hybrid algorithms with variable step sizes (PVSHM) designed for the efficient numerical solution of third-order initial value problems (IVPs) governed by Lane–Emden-type equations (LETEs). These equations frequently arise in various scientific and engineering fields, including astrophysics, fluid dynamics, and chemical engineering, where they serve as fundamental mathematical models for complex physical systems. The proposed methods utilize three internal interpolation points and employ specially designed starting formulas on the initial interval to properly handle the singular behavior at the origin. These formulations avoid the direct evaluation of the singular term while preserving the accuracy of the numerical scheme. To ensure both high accuracy and computational efficiency, the PVSHM operates within a variable step-size framework supported by an adaptive local error control strategy, which maintains truncation errors within predefined tolerances. Extensive numerical experiments validate the accuracy, stability, and computational effectiveness of the proposed algorithms, confirming their ability to efficiently solve a wide range of third-order LETE-based IVPs encountered in applied science and engineering contexts.

Keywords: Lane–Emden-type equations; Third-order initial value problems; Hybrid methods; Variable step-size algorithm; Adaptive error control

*Corresponding author

Received 30 December 2025; revised 20 June 2026; accepted 14 July 2026

Mehdi Fazli
Department of Mathematics, Ard.C., Islamic Azad University, Ardabil, Iran. e-mail: Me.Fazli@iauardabil.ac.ir

How to cite this article
Fazli, M., Adaptive hybrid numerical schemes with singularity treatment for third order Lane–Emden-type equations. *Iran. J. Numer. Anal. Optim.*, 2027; 17(1): 1-24. https://doi.org/10.22067/ijnao.2026.97038.1812

AMS subject classifications (2020): 65L05; 65L06; 65L20; 65L50

1 Introduction

Singular initial value problems (SIVPs) governed by Lane–Emden-type equations (LETEs) are pivotal in various scientific and engineering disciplines, including astrophysics, fluid dynamics, and chemical engineering. These equations often exhibit singular behavior at the origin, posing significant challenges for numerical solutions. Traditional methods may struggle with such singularities, necessitating the development of advanced numerical techniques.

Recent advancements have introduced hybrid methods with variable step sizes to address these challenges effectively. For instance, a pair of optimized Nyström methods with variable step sizes was proposed, demonstrating improved accuracy and efficiency in solving SIVPs of Lane–Emden–Fowler type equations. Similarly, a variable step-size formulation of a pair of block methods was developed, effectively solving SIVP models of Lane–Emden–Fowler type equations [20, 21].

Building upon these developments, this paper introduces two novel hybrid algorithms with variable step sizes designed specifically for efficiently solving SIVPs governed by LETEs. The proposed methods utilize three internal interpolation points and employ modified formulations on the initial subinterval to accommodate the singular behavior near the origin. Rather than eliminating the singularity, these starting formulas avoid the direct evaluation of the singular term at the origin and ensure accurate numerical approximations. To ensure both high accuracy and computational efficiency, the pair of variable-step-size hybrid methods (PVSHM) operates within a variable step-size framework supported by an adaptive local error control strategy, maintaining truncation errors within predefined tolerances.

Extensive numerical experiments validate the accuracy, stability, and computational efficiency of the proposed algorithms. The results demonstrate their capability to effectively solve a wide range of third-order LETE-based initial value problems (IVPs) encountered in applied science and engineering contexts.

Apart from their role in characterizing fluid motion in engineering systems and describing the behavior of stellar or astrophysical structures, the LETEs offer a comprehensive mathematical framework for analyzing a wide spectrum of physical processes. Numerous studies [8, 25, 24] have extensively documented the versatility of LETEs in various branches of engineering, physics, and applied sciences. Motivated by these diverse and significant applications, the present work is devoted to developing an efficient numerical approach for determining the function $z : [0, T] \to \mathbb{R}$ that satisfies the following third-order Lane–Emden-type model problem:

$$z'''(t) = l(t, z(t)) - k(t)z'(t) - \frac{\mu}{t}z''(t), \qquad t \in (0, T),$$
$$z(0) = z_0, \qquad z'(0) = z_0', \qquad z''(0) = z_0'', \tag{1}$$

where the parameters μ, z_0, z_0', z_0'', and $T > 0$ are known values and $k : [0,\ T] \to \mathbb{R}$ and $l : [0,\ T] \times \mathbb{R} \to \mathbb{R}$ are given continuous functions.

Given the broad range of real-world phenomena described by LETEs, the pursuit of analytical and numerical solutions has attracted significant attention from researchers in multiple scientific domains. The inherent nonlinearity and singularity at $t = 0$ further complicate the analytical treatment of these problems, making the derivation of closed-form solutions exceedingly difficult. As a result, numerical approaches have emerged as essential tools, enabling the reliable approximation of solutions and providing deeper insight into the complex behaviors governed by the singular structure of LETEs.

Among the various analytical approaches developed for solving differential equations such as the LETE presented in (1), a homotopy perturbation method has proven to be particularly effective. In this technique, a homotopy is formulated by introducing an auxiliary parameter that continuously deforms a simple problem into the original nonlinear one. The solution is then represented as a power series expansion in terms of the homotopy parameter. Substituting this expansion into the governing equation yields a sequence of recursive relations for the series coefficients, which can be solved iteratively to obtain progressively accurate approximations. Several other semi-analytical methods have also been employed to tackle (1) and its variants, including the Adomian decomposition method and the variational iteration method; detailed discussions can be found in [14, 32, 11, 2, 1, 15, 6, 17, 5].

The spline-based technique represents another numerical strategy for approximating the solutions of differential equations, including the LETE defined in (1). In this approach, the problem domain is divided into a set of subintervals, and the unknown solution is approximated within each subinterval using piecewise polynomial functions. The resulting system of algebraic equations is then solved, followed by an analysis of the convergence and stability of the approximation. The spline method offers several benefits, such as high accuracy and the capability to effectively manage nonuniformly spaced data. Nevertheless, its performance deteriorates when applied to stiff problems, and it often becomes computationally demanding for SIVPs because of the complexity associated with constructing spline functions and solving the corresponding algebraic systems.

Collocation techniques constitute another important category of numerical approaches employed for solving differential equations, including the third-order LETE presented in (1). In

this framework, the computational domain is partitioned into discrete nodes, and the original differential equation is approximated by a system of algebraic equations enforced at selected collocation points. The approximate solution is represented through suitable basis functions, whose unknown coefficients are determined by satisfying the governing equation at these points. By solving the resulting algebraic system, an approximate numerical solution can be obtained. Collocation methods are capable of providing accurate results for a wide range of differential equations; however, their accuracy and stability are strongly dependent on the choice of collocation points and basis functions. Moreover, the convergence characteristics of these methods may differ depending on the nature of the problem and the particular collocation scheme adopted. Further details on spline-based and collocation techniques can be found in [10, 9, 13, 31, 3].

Several other numerical techniques have been developed to address LETEs and their related forms, including the finite difference method, Haar wavelet method, Chebyshev neural network approach, hybrid block algorithms, and various other computational schemes. Comprehensive discussions and comparative analyses of these approaches can be found in [27, 30, 12, 7, 18, 22, 16, 23, 4, 19, 26, 28]. However, most of these existing methods have been implemented using a constant step-size framework, which tends to be inefficient for problems characterized by rapid variations in the solution. Consequently, in order to maintain numerical stability and acceptable accuracy, many researchers have restricted their computations to small integration intervals, thereby limiting the overall efficiency of such methods.

In this study, a PVSHM is developed for the numerical solution of third-order LETEs. The proposed framework combines adaptive error control with a singularity treatment strategy at the origin, enabling accurate and efficient computations over wider integration intervals. The method is designed to improve computational efficiency while maintaining the desired level of accuracy for a broad class of LETE problems.

The main contributions of this work can be summarized as follows:

1. A new PVSHM is developed for the direct numerical solution of third-order LETEs without reducing the original problem to a system of lower-order equations.

2. Special starting formulas are constructed for the initial interval to accurately accommodate the singular behavior near the origin while avoiding the direct evaluation of the singular term.

3. An embedded lower-order technique (LOT) is incorporated for local error estimation, enabling adaptive step-size selection without requiring additional function evaluations.

4. A frozen Jacobian implementation is employed within the Newton iteration process, reducing the computational cost associated with repeated Jacobian evaluations and matrix factorizations.

5. Numerical experiments demonstrate that the proposed PVSHM achieves high accuracy with fewer integration steps and reduced computational effort compared with several existing methods reported in the literature.

The manuscript is organized as follows. Section 2 presents the proposed numerical method, while Sections 3 and 4 provide a comprehensive theoretical analysis along with a strategy for estimating the local truncation error. Section 5 addresses the implementation details of the method, and Section 6 illustrates its performance through numerical examples. Finally, Section 7 summarizes the main findings and conclusions of the study.

2 Development of the proposed PVSHM

The polynomial $Q(t)$, defined as

$$z(t) \approx Q(t) = \sum_{n=0}^{7} b_n t^n, \tag{2}$$

is employed to approximate the analytical solution $z(t)$ at the grid points t_n, $n = 0, 1, \ldots, N$, such that $t_0 = 0$ and $t_N = T$ span the entire interval $[0, T]$. The selection of three internal interpolation points was motivated by the objective of constructing a high-order hybrid scheme while maintaining a manageable computational structure. Together with the interpolation and collocation conditions employed in the derivation, these points provide sufficient information to uniquely determine the coefficients of the seventh-degree polynomial approximation. This configuration leads to the eighth-order accuracy established later in the paper, while avoiding the substantial increase in algebraic complexity associated with additional internal points. Although higher-order approximations may be obtained by introducing more interpolation points, the resulting coefficient expressions and nonlinear systems become considerably more complicated, often with limited practical gains in efficiency. Therefore, the adopted configuration represents a balanced compromise between accuracy, stability, and computational cost.

The coefficients b_n in the polynomial can be determined by imposing collocation conditions at suitable points. To compute the unknown coefficients $b_0, b_1, \ldots, b_7$, three intermediate points $t_{n+p_1} = t_s + p_1 h$, $t_{n+p_2} = t_s + p_2 h$, and $t_{n+p_3} = t_s + p_3 h$ are considered within the interval

$[t_n, t_{n+1}]$, where $h = t_{j+1} - t_j$ denotes the step size and $0 < p_1 < p_2 < p_3 < 1$. By evaluating the approximation in (2) and its derivatives at these specific points, a system of equations is established to determine the coefficients b_n:

$$\begin{aligned} &Q(t_n) = z_n, && Q^{'}(t_n) = z_n^{'}, && Q^{''}(t_n) = z_n^{''}, \\ &Q^{'''}(t_n) = f_n, && Q^{'''}(t_{n+p_1}) = f_{n+p_1}, && Q^{'''}(t_{n+p_2}) = f_{n+p_2}, \\ &Q^{'''}(t_{n+p_3}) = f_{n+p_3}, && Q^{'''}(t_{n+1}) = f_{n+1}, \end{aligned}$$

where z_{n+j}, z'_{n+j}, z''_{n+j}, and f_{n+j} with $j = 0,\ p_1, p_2, p_3,\ 1$ denote approximate values of $z(t_{n+j})$, $z'(t_{n+j})$, $z''(t_{n+j})$ and $z'''(t_{n+j})$, respectively.

To derive the polynomial approximation $Q(t)$, the coefficients b_n $(n = 0, \ldots, 7)$ in (2) are expressed in terms of a local variable substitution $t = t_n + xh$. This transformation yields the following polynomial representation:

$$\begin{aligned} Q(t_n + xh) = {} & a_0(x)z_n + ha_1(x)z'_n + h^2 a_2(x)z''_n \\ & + h^3\big[\beta_0(x)f_n + \beta_{p_1}(x)f_{n+p_1} + \beta_{p_2}(x)f_{n+p_2} \\ & + \beta_{p_3}(x)f_{n+p_3} + \beta_1(x)f_{n+1}\big]. \end{aligned} \tag{3}$$

The coefficients $a_0,\ a_1,\ a_2,\ \beta_0,\ \beta_{p_1},\ \beta_{p_2},\ \beta_{p_3}$ and β_1, which depend on $p_1,\ p_2$, and p_3, are continuous functions of x. These coefficients are determined by the values of b_n and the intermediate collocation points employed in the approximation. Once these coefficients are obtained, the next step is to substitute them into (3) and evaluate $Q(t)$, $Q^{'}(t)$, and $Q^{''}(t)$ at $t_{n+1} = t_n + h$. This procedure yields the approximations for $z(t_{n+1})$, $z^{'}(t_{n+1})$ and $z^{''}(t_{n+1})$. For brevity, and without loss of generality, assuming $n = 0$, the resulting approximations are expressed as follows:

$$\begin{aligned} z_1 &= z_0 + hz_0^{'} + \frac{h^2}{2}z_0^{''} + \frac{h^3}{1680}\left[3\left(99f_{p_1} - 64f_{p_2} + 36f_{p_3}\right) + 72f_0 - 5f_1\right], \\ z_1^{'} &= z_0^{'} + hz_0^{''} + \frac{h^2}{120}\left[54f_{p_1} - 32f_{p_2} + 27f_{p_3} + 11f_0\right], \\ z_1^{''} &= z_0^{''} + \frac{h}{120}\left(81f_{p_1} - 64f_{p_2} + 81f_{p_3} + 11f_0 + 11f_1\right), \end{aligned} \tag{4}$$

where $p_1 = \frac{1}{3}$, $p_2 = \frac{1}{2}$, and $p_3 = \frac{3}{4}$. For the proposed PVSHM to be implemented in block mode, the evaluation of $Q(t)$, $Q'(t)$, and $Q''(t)$ at the intermediate points t_{n+p_1}, t_{n+p_2}, and t_{n+p_3} is carried out. This evaluation yields nine additional formulas, which are then incorporated with the formulas provided in (4) to construct the proposed method.

It is important to note that the formulas presented above, as well as the additional derived relations, cannot be directly applied to solve the LETE in (1) due to the presence of a singularity

at $t_0 = 0$. This singularity prevents the evaluation of $f_0 = f(t_0,\ z_0,\ z_0^{'}, z_0^{''})$, which is essential for the numerical solution of the LETE defined in (1). To overcome this issue and enhance accuracy near the origin, alternative formulas must be employed for the subinterval $[t_0, t_1]$. The construction of these modified formulas follows the same procedure used to derive (4), with the exception that the evaluation of f_0 is omitted. These specially designed formulas effectively eliminate the singularity at the origin and are expressed as follows:

$$\begin{aligned} z_1 &= z_0 + hz_0^{'} + \frac{h^2}{2} z_0^{''} + \frac{h^3}{240}\left(3\left(45f_{p_1} - 64f_{p_2} + 36f_{p_3}\right) - 11f_1\right), \\ z_1^{'} &= z_0^{'} + hz_0^{''} + \frac{h^2}{120}\left(153f_{p_1} - 208f_{p_2} + 126f_{p_3} - 11f_0\right), \\ z_1^{''} &= z_0^{''} + \frac{h}{2}\left(3f_{p_1} - 4f_{p_2} + 3f_{p_3}\right), \end{aligned} \tag{5}$$

where $p_1 = \frac{1}{3}$, $p_2 = \frac{1}{2}$, and $p_3 = \frac{2}{3}$. Additionally, the remaining formulas are obtained by evaluating $Q(t)$, $Q^{'}(t)$, and $Q^{''}(t)$ at the intermediate points t_{p_1}, t_{p_2}, and t_{p_3}. By combining the main formulas in (4) and (5) with these additional relations obtained from the evaluations of $Q(t)$, $Q^{'}(t)$, and $Q^{''}(t)$ at the intermediate points, a pair of hybrid methods is constructed.

3 Characteristics of the PVSHM

This section presents a detailed theoretical investigation of the proposed PVSHM method and examines its fundamental properties. The analysis focuses on determining the order of accuracy, assessing the stability characteristics, and deriving the local truncation error associated with the approximate solution.

3.1 Order of accuracy for the proposed PVSHM

To determine the order of accuracy of the proposed PVSHM, the set of equations presented in (4) is combined with nine additional relations. This combination yields the following expressions:

$$\overline{S}_1 z_n = h\overline{S}_2 z'_n + h^2\overline{S}_3 z''_n + h^3\overline{S}_4 F_n, \tag{6}$$

where $\overline{S}_1$, $\overline{S}_2$, $\overline{S}_3$, and $\overline{S}_4$ are coefficient matrices generated from (4) and additional formulas from the evaluation of $Q(t)$, $Q'(t)$, and $Q''(t)$ at the intermediate points t_{n+p_1}, t_{n+p_2}, t_{n+p_3} and

$$\mathbf{z}_n = (z_n, z_{n+p_1}, z_{n+p_2}, z_{n+p_3}, z_{n+1})^T,$$
$$\mathbf{z}_n^{'} = \left(z_n^{'}, z_{n+p_1}^{'}, z_{n+p_2}^{'}, z_{n+p_3}^{'}, z_{n+1}^{'}\right)^T,$$
$$\mathbf{z}_n^{''} = \left(z_n^{''}, z_{n+p_1}^{''}, z_{n+p_2}^{''}, z_{n+p_3}^{''}, z_{n+1}^{''}\right)^T,$$
$$\mathbf{F}_n = (f_n, f_{n+p_1}, f_{n+p_2}, f_{n+p_3}, f_{n+1})^T.$$

Assuming $z(t)$ is sufficiently differentiable, the operator $\mathcal{L}$ associated with (4) can be represented as

$$\begin{aligned}&\mathcal{L}(z(t);\ h) \\ &= \sum_{j\in I}\left[\psi_j z(t+jh) - h\sigma_j z^{'}(t+jh) - h^2\gamma_j z^{''}(t+jh) - h^3 v_j z^{'''}(t\ +\ jh)\right],\end{aligned} \tag{7}$$

where ψ_j, σ_j, γ_j, and v_j are vectors derived from the coefficient matrices $\bar{S}_1$, $\bar{S}_2$, $\bar{S}_3$ and $\bar{S}_4$, respectively. Assuming that $z(t)$ is sufficiently differentiable, the operator $\mathcal{L}$ can be expressed as a series expansion in powers of h using the following Taylor series:

$$\begin{aligned}\mathcal{L}\left(z(t);\ h\right) =& M_0 z(t) + M_1 h z^{'}(t) + M_2 h^2 z^{''}(t) + M_3 h^3 z^{'''(t)} + \cdots \\ &+ M_r h^r z^{r(t)} + \cdots,\end{aligned} \tag{8}$$

where M_r is given by

$$\begin{aligned}M_r =& \frac{1}{r!}\left[\sum_{j\in I} j^r\psi_j - r\sum_{j\in I} j^{r-1}\sigma_j\right. \\ &\left. -r(r-1)\sum_{j\in I} j^{r-2}\gamma_j - r(r-1)(t-2)\sum_{j\in I} j^{r-3}v_j\right],\end{aligned} \tag{9}$$

For $r = 0, 1, 2, \ldots$ and $I = \{0,\ p_1,\ p_2,\ p_3, 1\}$, by applying the equations in (4) to (8) and (9), the order (p) of the main formulas and their corresponding local truncation errors $(\mathcal{L})$ can be expressed as follows:

$$\begin{aligned}&P = 5, \qquad \mathcal{L}\left(z_{n+1}; h\right) = \frac{h^8}{362\ 880}\ z^{(8)}(t_n) + O(h^9), \\ &P = 5, \qquad \mathcal{L}\ \left(z_{n+1}^{'}; h\right) = \frac{h^8}{181\ 440}\ z^{(7)}(t_n) + O(h^8), \\ &P = 6, \qquad \mathcal{L}\ \left(z_{n+1}^{''}; h\right) = \frac{h^9}{1\ 088\ 640}\ z^{(7)}(t_n) + O(h^{10}).\end{aligned} \tag{10}$$

Assessing the consistency of a numerical method requires a thorough examination of the order of its approximation formulas. Since the proposed method possesses an order higher than one, it is considered consistent and is capable of providing accurate approximations for the LETE in (1). It is important to note, however, that while consistency is a necessary condition for convergence, it does not by itself guarantee the convergence of a numerical method. Therefore, additional theoretical properties of the proposed PVSHM, such as zero-stability, must be carefully analyzed to ensure the method's convergence.

3.2 Zero-stability for the proposed PVSHM

As h tends to zero, the behavior of (6) is closely related to the zero-stability of the proposed method. This behavior can be represented more succinctly by the following equation: $B_1^{(0)}\ \bar{\mathbf{Z}}_\mu - B_1^{(1)}\ \bar{\mathbf{Z}}_{\mu-1} = 0$, where

$$\bar{\mathbf{Z}}_\mu = \left(z_{n+1}, z_{n+p_3}, z_{n+p_2}, z_{n+p_1}\right)^T,$$

$$\bar{\mathbf{Z}}_{\mu-1} = \left(z_n, z_{n+p_3-1}, z_{n+p_2-1}, z_{n+p_1-1}\right)^T,$$

$$B_1^{(0)} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad B_1^{(1)} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

It should be noted that the above simplified form is obtained by taking the limit as $h \to 0$, providing a reliable assessment of the proposed method's accuracy in approximating the true solution of the problem. For the PVSHM to be considered zero-stable, its primary characteristic polynomial $P(s)$ must satisfy specific root conditions. These conditions require that the roots s_j obey $|s_j| \leq 1$, and any roots with $|s_j| = 1$ must have multiplicities not exceeding two. For the proposed PVSHM, the characteristic polynomial is given by $P(s) = (s-1)s^3$, which satisfies all the criteria for zero-stability. Consequently, the PVSHM is both consistent and zero-stable, fulfilling the necessary conditions for convergence. Therefore, the proposed PVSHM is a convergent numerical method.

4 Strategy to estimate the local error

This section introduces a variable step-size formulation for the PVSHM, which offers several advantages. In particular, this approach allows the use of larger integration intervals during computation, thereby increasing computational efficiency. To ensure accurate and reliable solutions of the LETE in (1), the truncation error is maintained below a specified relative tolerance. By employing an embedding technique, the proposed method can be implemented in an adaptive mode, which requires the construction of an additional LOT. The LOT is essential for accurately estimating the local error of the PVSHM. Notably, the LOT is derived using the same function evaluations as the PVSHM, so no additional computational cost is incurred. For the accurate estimation of the local error of the proposed PVSHM, the following LOT is utilized:

$$z^*_{n+1} = 2z_n - 9z_{n+p_1} + 8z_{n+p_2} + h^3\left(\frac{f_n}{1728} + \frac{f_{n+p_1}}{32} + \frac{f_{n+p_2}}{216} + \frac{11f_{n+p_3}}{576}\right) \tag{11}$$

with $\mathcal{L}\left(z^*_{n+1}\right) = \frac{17h^7 z^{(7)}(t_n)}{11{,}197{,}440} + O(h^8)$.

This LOT involves the computation of z^*_{n+1} using a combination of z_n, z_{n+p_1}, z_{n+p_2} and the corresponding function evaluations f_n, f_{n+p_1}, f_{n+p_2}, and f_{n+p_3}, providing an accurate estimate of the local error. Specifically, z^*_{n+1} is computed as

$$z^*_{n+1} = 2z_n - 9z_{n+p_1} + 8z_{n+p_2} + h^3(1728f_n + 32f_{n+p_1} + 216f_{n+p_2} + 576f_{n+p_3}).$$

The local error estimate, denoted by ERRES, is then defined as

$$\text{ERRES} = \frac{\parallel z^*_{n+1} - z_{n+1} \parallel}{1+ \parallel z_{n+1} \parallel}.$$

This formulation allows for the adaptive selection of the step size in the proposed PVSHM, ensuring that the local truncation error remains within the desired tolerance.

The formula above measures the difference between the approximations z^*_{n+1} and z_{n+1}, which are obtained using the LOT in (11) and the PVSHM in (6), respectively. The computed approximation is considered acceptable if the absolute value of ERRES is less than or equal to the specified relative error tolerance (RET).

Conversely, if the absolute value of ERRES exceeds the RET, the computed result is rejected, and the step size must be adjusted to improve the accuracy of the approximation. This adjustment is performed using the following procedure:

The adaptive step size is updated according to the formula

$$h_{\text{var}} = \Upsilon \; h_{\text{old}} \left(\frac{\text{RET}}{\| \text{ ERRES } \|} \right)^{\frac{1}{p+3}},$$

where $\Upsilon = 0.85$ is a suitable adjustment factor, and p denotes the order of the LOT in (11). This adaptive strategy for modifying the step size is particularly important because different problems may require different levels of accuracy. By appropriately adjusting the step size, the LETE in (1) can be solved efficiently, achieving the desired accuracy while minimizing computational cost.

5 Implementation

To implement the PVSHM, the error estimation procedure presented in Section 4 is employed. This requires reformulating the formulas given in (4) and (5), along with the additional relations obtained by evaluating $Q(t)$, $Q^{'}(t)$, and $Q^{''}(t)$ at the intermediate points, into the form $F(\widetilde{\mathbf{Z}}) = 0$. Here, the vector $\widetilde{\mathbf{Z}}$ is used to represent the unknowns in this reformulation process,

$$\widetilde{\mathbf{Z}} = \; (z_{p_1}, \; z^{'}{}_{p_1}, z^{''}{}_{p_1}, \; z_{p_2}, \; z^{'}{}_{p_2}, z^{''}{}_{p_2}, \; z_{p_3}, \; z^{'}{}_{p_3}, z^{''}{}_{p_3}, \; z_1, \; z^{\prime}{}_1, \; z^{\prime\prime}{}_1).$$

By representing the unknowns as described above, the computations for the PVSHM can be performed efficiently. The resulting nonlinear equations can be solved using Newton's method, which iteratively updates the approximation according to

$$\widetilde{\mathbf{Z}}^{(j+1)} = \widetilde{\mathbf{Z}}^{(j)} - (J_0)^{-1} F^{(j)},$$

where J_0 denotes the Jacobian matrix of $F(\widetilde{\mathbf{Z}})$ and the superscript (j) indicates the current iteration. Here, $\widetilde{\mathbf{Z}}^{(j)}$ represents the approximation at iteration j, and $F^{(j)}$ is a vector that depends on the current approximation. The frozen Jacobian matrix of F is denoted by J_0.

In the present implementation, a frozen Jacobian strategy is adopted, where the Jacobian matrix is evaluated once at the initial approximation and subsequently reused throughout the Newton iterations. This approach avoids the repeated computation and factorization of the Jacobian matrix at every iteration, thereby reducing the overall computational cost. Since the nonlinear systems generated by the PVSHM are solved using sufficiently accurate initial predictors, the Jacobian matrix typically varies only mildly between successive iterations. Consequently, the frozen Jacobian strategy provides an efficient compromise between computational efficiency and convergence performance.

To apply Newton's method for solving the nonlinear system, suitable initial values for the approximation must be provided. These initial values are obtained using the following Taylor

series formulas:

$$z_{n+j} = z_n + (jh)z'_n + \frac{(jh)^2}{2}z''_n + \frac{(jh)^3}{6}f_n$$

$$z'_{n+j} = z'_n + (jh)z''_n + \frac{(jh)^2}{2}f_n$$

$$z''_{n+j} = z''_n + (jh)f_n,$$

for $n = 1, 2, \ldots, N-1$, and $j = p_1, p_2, p_3,\ 1$.

6 Numerical examples

This section presents numerical examples to validate the performance and demonstrate the efficacy of the proposed PVSHM in solving the LETE in (1). Additionally, the results obtained using PVSHM are compared with those from various numerical methods reported in the scientific literature. The maximum absolute error (Err) is used to assess the accuracy of the numerical solutions. This measure allows for direct comparison between the computed solution and both exact and existing reference solutions. The Err is defined as

$$\text{ERR} = \max_{j=0,1,\ldots,N} \parallel z(t_j) - z_j \parallel,$$

where $z(t_j)$ is the exact solution, while z_j represents a numerical approximation. Using this formula, one can quantitatively assess the accuracy of PVSHM and juxtapose its performance with that of other methods in the available literature.

To facilitate the presentation of results and enhance readability, several abbreviations are introduced. PVSHM denotes the newly proposed method. HBM refers to the hybrid block method. CBST stands for the cubic B-spline technique. KHST denotes the kernel Hilbert space technique. PBM represents the pair block method.

Several computational variables are also used throughout the numerical experiments. CT measures the computational time, in seconds, required to solve a given LETE problem. h_0 denotes the initial step size used in the numerical solution. NoS represents the total number of steps employed during computation, and RET specifies the relative error tolerance.

The benchmark results reported for CBST, KHST, HBM, and PBM were taken directly from the corresponding references cited in Tables 1–5. Consequently, the computational settings, implementation details, numerical precision, and hardware specifications associated with these methods are those reported in the original publications. The purpose of these comparisons is to provide a qualitative assessment of the accuracy and efficiency of the proposed PVSHM relative to

existing approaches. Since implementation environments may differ among studies, the reported comparisons should be interpreted within this context.

Example 1. Consider the studied LETE, which is given by

$$z^{'''}(t) = -\frac{2}{t} + \frac{9(t^6+8)}{8z(t)^5}, \quad z(0) = 1, \quad z^{'}(0) = 0, \qquad z^{''}(0) = 0, \quad t \in [0, t_N], \tag{12}$$

with the exact solution $z(t) = \sqrt{1+t^3}$. The results of applying different numerical methods to Numerical Example 1 are summarized in Table 1. These results highlight the effectiveness of PVSHM, CBST, and KHST in approximating the solution of Numerical Example 1. For the parameters RET $= 10^{-7}$ and $h_0 = 10^{-1}$, the proposed PVSHM achieves Err $= 8.9681 \times 10^{-8}$ using only six steps, demonstrating its high efficiency. When the parameters are tightened to RET $= 10^{-11}$ and $h_0 = 10^{-3}$, PVSHM delivers an even more accurate result, with Err $= 5.5020 \times 10^{-10}$ using 14 steps, emphasizing the method's superior convergence properties. In comparison, while the CBST and KHST methods also provide reasonable approximations, they exhibit larger Errs under both parameter sets. Specifically, the CBST method produces Errs of 1.0×10^{-7} and 5.1900×10^{-7} with 10 and 40 steps, respectively. Similarly, the KHST method yields Errs of 7.7839×10^{-6} and 4.3964×10^{-6} using 20 and 35 steps, respectively. These comparisons demonstrate that the PVSHM not only achieves higher accuracy but also requires fewer computational steps compared to other methods.

Table 1: Comparison of the results obtained for Numerical Example 1 over the interval $t \in [0, 1]$.

Method	NoS	Err
$PVSHM$	6	8.9681×10^{-8}
$CBST$	10	5.2500×10^{-5}
$KHST$	20	7.7839×10^{-6}
$PVSHM$	14	5.5020×10^{-10}
$CBST$	40	5.1900×10^{-7}
$KHST$	35	4.3964×10^{-6}

The data presented in Table 1 clearly indicate that the PVSHM achieves the smallest Errs among the three compared methods for both sets of parameters. This demonstrates that the PVSHM consistently outperforms the CBST and KHST methods in terms of accuracy. The superior performance of PVSHM further highlights its robust convergence when applied to the LETE defined in (12). Additional evidence of this performance is provided in Figure 1, where the analytical and PVSHM solutions are depicted on the left, and the corresponding absolute errors are shown on the right. To generate these plots, the computations were performed using RET $= 10^{-12}$, an initial step size of $h_0 = 10^{-3}$, and the interval $t \in [0, 50]$.

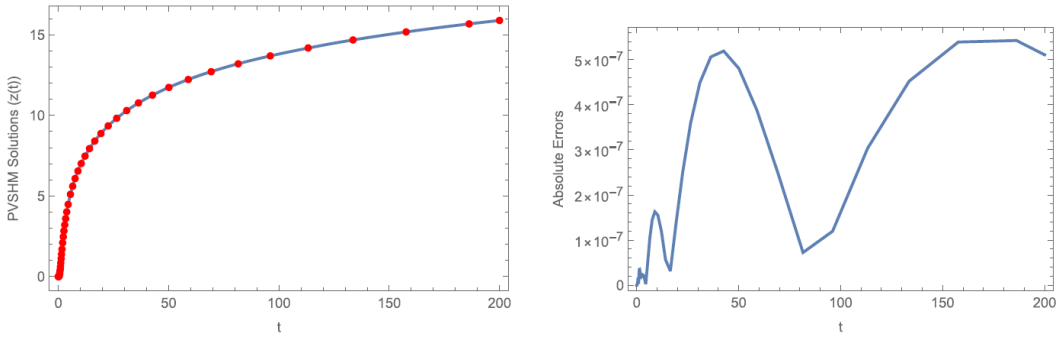


Figure 1: Absolute errors, analytical solution, and PVSHM approximation for Numerical Example 1 obtained with RET $= 10^{-12}$ and initial step size $h_0 = 10^{-3}$ over the interval $t \in [0, 50]$.

Example 2. Consider the LETE, which is given by

$$z'''(t) = 6(t^6 + 2t^3 + 10)e^{-3z(t)} + \frac{6}{t^2}z'(t) + \frac{6}{t}z''(t),$$
$$z(0) = 0, \quad z'(0) = 0, \quad z''(0) = 0, \quad t \in [0, t_N]. \tag{13}$$

The analytical solution to this LETE is

$$z(t) = \ln(1 + t^3).$$

Table 2 compares the Errs for Numerical Example 2 obtained using the PVSHM, CBST, and KHST methods. The results clearly indicate that the PVSHM consistently outperforms both the CBST and KHST methods in terms of accuracy. In the first set of parameters, with RET $= 10^{-7}$ and $h_0 = 10^{-2}$, the PVSHM requires only seven steps to achieve an error of 9.8681×10^{-8}. This error is significantly smaller than those obtained using the CBST and KHST methods, which yield errors of 5.1500×10^{-6} and 2.6839×10^{-6}, respectively, while requiring a larger NoS values (20 and 30). The smaller error achieved by the PVSHM with fewer steps demonstrates its superior efficiency and accuracy compared to the CBST and KHST methods.

In the subsequent set of parameters using RET $= 10^{-7}$ and $h_0 = 10^{-2}$ (as shown in Table 2), the PVSHM method further demonstrates its superior performance by achieving an error of 3.4020×10^{-9} with only 11 steps. In contrast, the CBST and KHST methods require 40 and 45 steps, respectively, and yield higher errors of 4.1800×10^{-7} and 1.2964×10^{-6}, indicating less accurate approximations compared to the PVSHM. Figure 2 provides a graphical comparison of these results: The absolute errors are shown on the right, while the analytical and PVSHM

solutions are plotted on the left for Numerical Example 2. The computations for Figure 2 were performed over the interval $t \in [0, 200]$.

Table 2: Comparison of the results obtained for Numerical Example 2 over the interval $t \in [0, 1]$.

Method	NoS	Err
$PVSHM$	7	9.8681×10^{-8}
$CBST$	20	5.1500×10^{-6}
$KHST$	30	2.6839×10^{-6}
$PVSHM$	11	3.4020×10^{-9}
$CBST$	40	4.1800×10^{-7}
$KHST$	45	1.2964×10^{-6}

Example 3. Consider the LETE previously solved in [29], given by

$$\begin{aligned} z'''(t) &= -\frac{4}{t}z''(t) - z(t)^m, \\ z(0) = 1, \quad z'(0) = 0, \quad z''(0) = 0, \quad t \in [0, t_N], \end{aligned} \tag{14}$$

where $m \geq 0$ is a constant. For $m = 0$, the exact solution of (14) is $z(t) = 1 - \frac{t^3}{30}$, and for $m \geq 1$, the exact solution is unknown.

For the computation of the error in Numerical Example 3 with $m = 5$, the NDSolve solution is employed as the reference solution. Numerical solutions are obtained using three different relative tolerances. Table 3 presents the results, clearly demonstrating that the PVSHM method outperforms the HBM in terms of accuracy. Furthermore, the effectiveness of the PVSHM method is illustrated in Figure 3, where the absolute error plots are displayed on the right, while the analytical solution (for $m = 0$) and the PVSHM solution are shown on the left for Numerical Example 3.

It is worth noting that the singular behavior of (14) is governed by the parameter μ, which directly controls the strength of the singular term $\left(\frac{\mu}{t}z''(t)\right)$. The proposed PVSHM employs specially designed starting formulas on the initial interval that avoid the direct evaluation of the singular term at the origin. Consequently, the effectiveness of the singularity treatment does not depend on a particular value of μ.

The numerical results obtained for Numerical Example 3 demonstrate that the proposed method remains stable and highly accurate in the presence of singular behavior. Moreover, since the starting formulas are derived independently of any specific value of the singularity parameter, the same treatment can be applied to a broader class of LETEs exhibiting different singularity strengths.

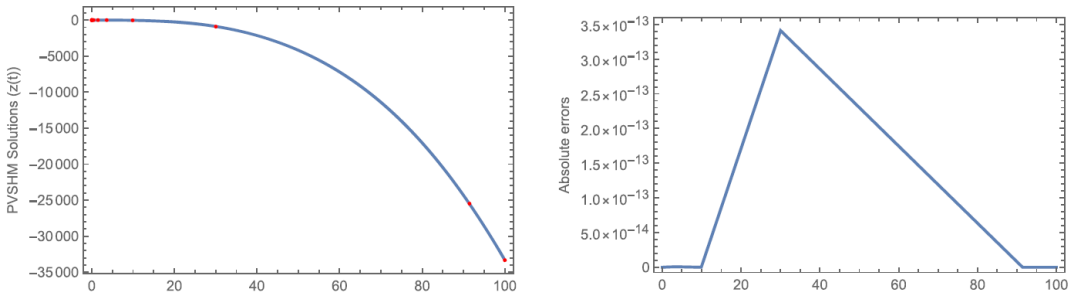


Figure 2: Comparison of the analytical solution and PVSHM approximation for Numerical Example 2, along with the corresponding absolute errors, computed using RET $= 10^{-8}$ and $h_0 = 10^{-3}$ over the interval $t \in [0, 200]$.

Table 3: Numerical comparison of the results for Numerical Example 3 over the interval $t \in [0, 10]$.

RET	Method	NoS	CT	Err
10^{-9}	PVSHM	74	0.0940	1.5269×10^{-6}
10^{-9}	HBM	111	0.1705	2.0191×10^{-6}
10^{-11}	PVSHM	143	0.1730	2.3768×10^{-8}
10^{-11}	HBM	201	0.2870	7.3995×10^{-8}
10^{-13}	PVSHM	275	0.3382	5.8712×10^{-10}
10^{-13}	HBM	375	0.4667	1.5499×10^{-9}

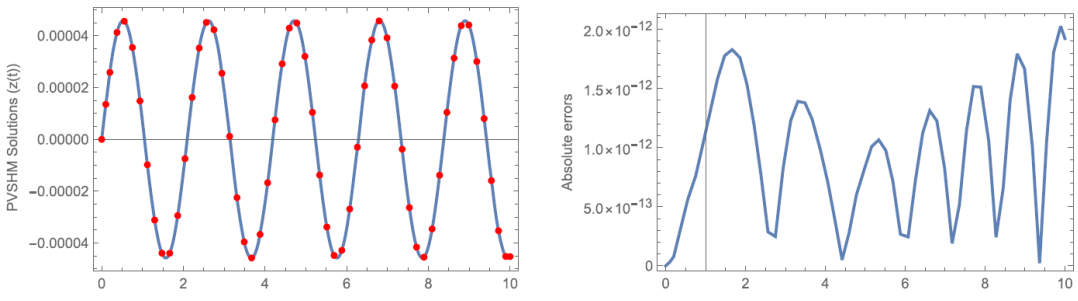


Figure 3: Comparison of the analytical solution and PVSHM approximation for Numerical Example 3, along with the corresponding absolute errors, computed for $m = 0$ using RET $= 10^{-10}$ and $h_0 = 10^{-2}$ over the interval $t \in [0, 100]$.

Example 4. Consider the following real-life problem in nonlinear science and engineering, specifically the self-adjoint LETE model, as discussed in [17]:

$$\varepsilon z'''(t) + \frac{1}{t} z''(t) + z(t) = 3\varepsilon \left[-27\varepsilon \cos(3t) - \frac{9 \sin(3t)}{t} + \sin(3t) \right], \tag{15}$$

where the initial conditions for Numerical Example 4 are obtained from the exact solution $z(t) = 3\varepsilon \sin(3t)$.

The results presented in Table 4 clearly indicate that the PVSHM method outperforms the PBM method in terms of the NoS, computation time (CT), and Err for all tested tolerances ($10^{-8}, 10^{-10}$, and 10^{-12}). This superiority is consistently observed across all considered tolerances. For instance, at a tolerance of 10^{-8}, the PVSHM method required 16 steps with a computation time of 0.0170 s and achieved a maximum absolute error of 1.5268×10^{-9}, whereas the PBM method required 32 steps, a computation time of 0.0458 s, and yielded a maximum absolute error of 1.0190×10^{-9}.

Table 4: Numerical comparison of the results for Numerical Example 4 over the interval $t \in [0, 10]$.

RET	Method	NoS	CT	Err
10^{-8}	PVSHM	16	0.0170	1.5268×10^{-9}
10^{-8}	PBM	32	0.0458	1.0190×10^{-9}
10^{-10}	PVSHM	30	0.0440	1.3767×10^{-11}
10^{-10}	PBM	58	0.0974	5.3994×10^{-11}
10^{-12}	PVSHM	55	0.0678	2.8711×10^{-12}
10^{-12}	PBM	124	0.1298	3.5498×10^{-12}

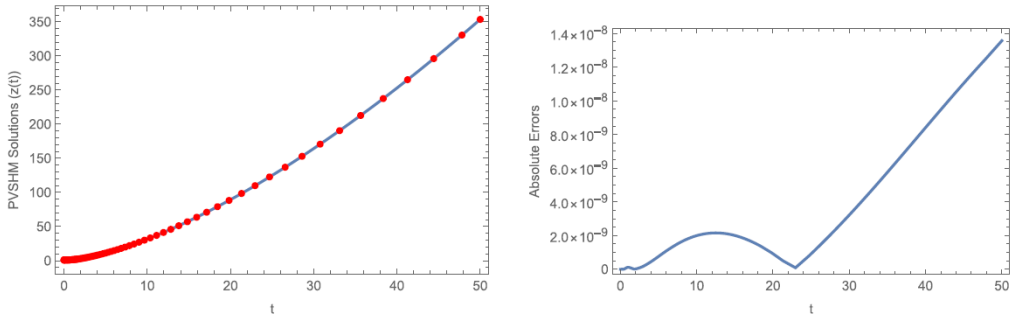


Figure 4: Comparison of the analytical solution and PVSHM approximation for (15), along with the corresponding absolute errors, computed using RET $= 10^{-12}$, $h_0 = 10^{-2}$, and $\varepsilon = 2^{16}$ over the interval $t \in [0, 10]$.

Furthermore, the superior performance of the PVSHM method is reinforced by the graphical representation shown in Figure 4. In this figure, the absolute errors are plotted on the right, while the analytical solution and the PVSHM approximation are displayed on the left for Numerical Example 4. Table 6 reports the efficiency-index comparison of the proposed PVSHM and the existing methods. For Numerical Examples 1 and 2, the efficiency index is computed using the

Table 5: Numerical comparison of the results for Numerical Example 5 over the interval $t \in [0, 10\pi]$.

RET	Method	NoS	CT	Err
10^{-9}	$PVSHM$	42	0.0941	2.5268×10^{-5}
10^{-9}	HBM	72	2.1704	1.0190×10^{-4}
10^{-10}	$PVSHM$	57	0.1720	6.3767×10^{-6}
10^{-10}	HBM	96	2.2860	1.3994×10^{-5}
10^{-11}	$PVSHM$	63	0.3372	1.8711×10^{-8}
10^{-11}	HBM	130	3.4657	5.5498×10^{-7}

number of integration steps, whereas for Numerical Examples 3–5 it is computed using the CT. Larger EI values indicate greater computational efficiency.

Table 6: Efficiency-index comparison of the proposed PVSHM and existing methods for the numerical examples considered in this study.

Example	Method	Error	NoS/CT	EI
Num. Ex. 1	PVSHM	8.9681×10^{-8}	NoS = 6	1.175
Num. Ex. 1	CBST	5.2500×10^{-5}	NoS = 10	0.428
Num. Ex. 1	KHST	7.7839×10^{-6}	NoS = 20	0.255
Num. Ex. 2	PVSHM	9.8681×10^{-8}	NoS = 7	1.001
Num. Ex. 2	CBST	5.1500×10^{-6}	NoS = 20	0.264
Num. Ex. 2	KHST	2.6839×10^{-6}	NoS = 30	0.186
Num. Ex. 3, 10^{-9}	PVSHM	1.5269×10^{-6}	CT = 0.0940	61.70
Num. Ex. 3, 10^{-9}	HBM	2.0191×10^{-6}	CT = 0.1705	33.55
Num. Ex. 4, 10^{-8}	PVSHM	1.5268×10^{-9}	CT = 0.0170	523.60
Num. Ex. 4, 10^{-8}	PBM	1.0190×10^{-9}	CT = 0.0458	196.10
Num. Ex. 5, 10^{-9}	PVSHM	2.5268×10^{-5}	CT = 0.0941	48.77
Num. Ex. 5, 10^{-9}	HBM	1.0190×10^{-4}	CT = 2.1704	18.40

Example 5. The final numerical example considers a system of LETEs from [21], given by

$$
\begin{aligned}
&z_1''' + \frac{z_1''}{t} - \frac{z_1'}{t^2} - z_2^2\left(2z_1z_2z_1' + 3(z_1^2+1)z_2'\right) = 0,\\
&z_2''' + \frac{3z_2''}{t} - \frac{3z_2'}{t^2} - z_2^4\left(2z_1z_2z_1' + 5(z_1^2+3)z_2'\right) = 0,\\
&z_3''' - \frac{3(z_1' - tz_3'')}{t^2} - 4t^3z_2^7 - 7t^4z_2^6z_2' + 5z_2^4(z_3^2-6)z_2' + 2z_2^5(6z_2' + z_3z_3') = 0, \quad t \in [0, t_N]. \quad (16)
\end{aligned}
$$

and the initial conditions are given by

$$z_1(0) = 1, \qquad z^{'}{}_1(0) = 0, \qquad z^{''}{}_1(0) = 1,$$

$$z_2(0) = 1, \qquad z'_2(0) = 0, \qquad z''_2(0) = -1,$$

$$z_3(0) = 1, \qquad z'_3(0) = 0, \qquad z''_3(0) = 1,$$

$$z'_n = (z'_n,\ z'_{n+p_1},\ z'_{n+p_2}, z'_{n+p_3},\ z'_{n+1})^T,$$

$$z''_n = (z''_n,\ \ z''_{n+p_1},\ z''_{n+p_2},\ z''_{n+p_3}, z''_{n+1}\ \ )^T,$$

$$F_n = (f_n,\ f_{n+p_1},\ f_{n+p_2},\ f_{n+p_3},\ f_{n+1})^T.$$

The analytical solution for this system is

$$z_1(t) = \sqrt{1+t^2}, \quad z_2(t) = \frac{1}{\sqrt{1+t^2}}, \quad z_3(t) = 1 - \frac{1}{\sqrt{1+t^2}}.$$

The results of Numerical Example 5 were compared using two numerical methods, PVSHM and HBM, over the time interval $t \in [0, 10\pi]$, as summarized in Table 5. The data clearly demonstrate that the PVSHM method consistently outperforms the HBM method across all tested RET levels of RETs. These results highlight the high accuracy of PVSHM, providing precise numerical solutions with minimal error. Combined with its efficiency in computational resources and execution time, the PVSHM method proves to be a reliable and effective approach for solving this system of LETEs.

6.1 Limitations and scope of applicability

Although the proposed PVSHM has demonstrated excellent accuracy and efficiency for the class of third-order LETEs considered in this study, several limitations should be noted. First, the method has been specifically developed for LETEs with singular behavior of the form considered in (1), and its performance for more general classes of singular differential equations has not been investigated. Second, the nonlinear systems arising from the discretization are solved using Newton iterations, and therefore the overall efficiency may depend on the quality of the initial approximation and the convergence behavior of the iterative solver. Third, while the proposed starting formulas effectively treat the singularity at the origin, additional modifications may be required for problems involving different singular structures or multiple singular points. Finally, the present work does not include a dedicated investigation of highly stiff problems, and therefore the performance of PVSHM for extremely stiff systems remains a topic for future research. These aspects provide promising directions for further development and extension of the proposed methodology. The present study focuses on the development, theoretical analysis, and

numerical validation of a specific PVSHM configuration. Consequently, a systematic sensitivity analysis with respect to algorithmic parameters, such as the adaptive step-size control factor, the embedded LOT, and the number of internal interpolation points, has not been undertaken. Although the numerical results indicate stable and reliable performance for the parameter settings adopted in this work, a comprehensive investigation of parameter sensitivity may provide additional insight into the robustness and optimization of the method. Such an analysis constitutes an interesting direction for future research. The proposed PVSHM uses an adaptive step-size mechanism that automatically adjusts based on the local behavior of the solution. When the solution changes rapidly and the error increases, the step size is reduced to maintain accuracy. In smoother regions, larger steps are used to improve efficiency. This makes the method robust for problems with moderate gradients and varying scales.

However, the study does not specifically examine highly oscillatory or very stiff problems. In such cases, more advanced techniques such as stiffness-aware step-size control, implicit methods, or oscillation-preserving strategies may be needed. Therefore, the performance of PVSHM under extreme stiffness or strong oscillations remains an open topic for future research.

7 Conclusion

A PVSHM has been presented for solving third-order LETEs. The proposed formulation incorporates a singularity treatment near the origin and an adaptive error-control mechanism for step-size selection. Theoretical analysis confirms the consistency, zero-stability, and convergence of the method. Numerical experiments demonstrate that the proposed approach produces accurate solutions while requiring fewer computational steps than the methods considered for comparison. Although the present study focuses on third-order LETEs, the underlying ideas of the proposed PVSHM, including the adaptive step-size strategy, embedded error estimation mechanism, and specially designed treatment of singular behavior near the origin, are not restricted solely to this class of problems. The methodology has the potential to be extended to other singular ordinary differential equations arising in astrophysics, plasma physics, heat-transfer models in spherical geometries, nonlinear reaction–diffusion systems, and various engineering applications involving singular IVPs. Consequently, the proposed framework may serve as a useful basis for the development of efficient numerical solvers for a broader class of singular differential equations. These results indicate that PVSHM is a reliable numerical technique for a wide range of third-order LETE models.

Funding

This research received no external funding.

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] Abdel-Salam, E.A.-B., Nouh, M.I. and Elkholy, E.A., *Analytical solution to the conformable fractional Lane-Emden type equations arising in astrophysics*, Sci. Afr. 8 (2020) e00386.

[2] Adomian, G., Rach, R. and Shawagfeh, N.T., *On the analytic solution of the Lane-Emden equation*, Found. Phys. Lett. 8(2) (1995) 161–181.

[3] Aminikhah, H. and Kazemi, S., *On the numerical solution of singular Lane-Emden type equations using cubic B-spline approximation*, Int. J. Appl. Comput. Math. 3(2) (2017) 703–712.

[4] Anastassi, Z.A., Kosti, A.A. and Rufai, M.A., *A parametric method optimised for the solution of the (2+1)-dimensional nonlinear Schrödinger equation*, Mathematics 11(3) (2023) 609.

[5] Azar, E.A., Jalili, B., Azar, A.A., Jalili, P., Atazadeh, M. and Ganji, D.D., *An exact analytical solution of the Emden-Chandrasekhar equation for self-gravitating isothermal gas spheres in the theory of stellar structures*, Phys. Dark Universe 42 (2023) 101309.

[6] Dehghan, M. and Tatari, M., *Finding approximate solutions for a class of third-order nonlinear boundary value problems via the decomposition method of Adomian*, Int. J. Comput. Math. 87(6) (2010) 1256–1263.

[7] Ghergu, M. and Rădulescu, V.D., *Bifurcation and asymptotics for the Lane-Emden-Fowler equation*, C. R. Math. 337(4) (2003) 259–264.

[8] Horedt, G.P., *Polytropes: applications in astrophysics and related fields*, Springer, 2004.

[9] Khuri, S. and Sayfy, A., *Numerical solution for the nonlinear Emden-Fowler type equations by a fourth-order adaptive method*, Int. J. Comput. Methods 11(01) (2014) 1350052.

[10] Lakestani, M. and Dehghan, M., *Four techniques based on the B-spline expansion and the collocation approach for the numerical solution of the Lane-Emden equation*, Math. Methods Appl. Sci. 36(16) (2013) 2243–2253.

[11] Liao, S., *A new analytic algorithm of Lane-Emden type equations*, Appl. Math. Comput. 142(1) (2003) 1–16.

[12] Mall, S. and Chakraverty, S., *Numerical solution of nonlinear singular initial value problems of Emden-Fowler type using Chebyshev Neural Network method*, Neurocomputing 149 (2015) 975–982.

[13] Mohammadzadeh, R., Lakestani, M. and Dehghan, M., *Collocation method for the numerical solutions of Lane-Emden type equations using cubic Hermite spline functions*, Math. Methods Appl. Sci. 37(9) (2014) 1303–1717.

[14] Nazari-Golshan, A., Nourazar, S.S., Ghafoori-Fard, H., Yildirim, A. and Campo, A., *A modified homotopy perturbation method coupled with the Fourier transform for nonlinear and singular Lane-Emden equations*, Appl. Math. Lett. 26(10) (2013) 1018–1025.

[15] Parand, K., Dehghan, M., Rezaei, A. and Ghaderi, S., *An approximation algorithm for the solution of the nonlinear Lane-Emden type equations arising in astrophysics using Hermite functions collocation method*, Comput. Phys. Commun. 181(6) (2010) 1096–1108.

[16] Parand, K., Shahini, M. and Dehghan, M., *Rational Legendre pseudospectral approach for solving nonlinear differential equations of Lane-Emden type*, J. Comput. Phys. 228(23) (2009) 8830–8840.

[17] Rufai, M.A. and Ramos, H., *Numerical integration of third-order singular boundary-value problems of Emden-Fowler type using hybrid block techniques*, Commun. Nonlinear Sci. Numer. Simul. 105 (2022) 106069.

[18] Rufai, M.A. and Ramos, H., *Numerical solution of Bratu's and related problems using a third derivative hybrid block method*, Comput. Appl. Math. 39(4) (2020) 322.

[19] Rufai, M.A. and Ramos, H., *Numerical solution of third-order boundary value problems by using a two-step hybrid block method with a fourth derivative*, Comput. Math. Methods 3(6) (2021) e1166.

[20] Rufai, M.A. and Ramos, H., *Solving SIVPs of Lane-Emden-Fowler type using a pair of optimized Nystrom methods with a variable step size*, Mathematics 11(6) (2023) 1535.

[21] Rufai, M.A. and Ramos, H., *Solving third-order Lane-Emden-Fowler equations using a variable stepsize formulation of a pair of block methods*, J. Comput. Appl. Math. 420 (2023) 114776.

[22] Rufai, M.A., *An efficient third-derivative hybrid block method for the solution of second-order BVPs*, Mathematics 10(19) (2022) 3692.

[23] Rufai, M.A., Tran, T. and Anastassi, Z.A., *A variable step-size implementation of the hybrid Nystrom method for integrating Hamiltonian and stiff differential systems*, Comput. Appl. Math. 42(4) (2023) 156.

[24] Sabir, Z., Raja, M.A.Z., Umar, M. and Shoaib, M., *Design of neuro-swarming-based heuristics to solve the third-order nonlinear multi-singular Emden-Fowler equation*, Eur. Phys. J. Plus 135(5) (2020) 410.

[25] Sabirov, R.K., *Solutions of the Lane-Emden and Thomas-Fermi Equations*, Russ. Phys. J. 45(2) (2002) 129–132.

[26] Shahni, J., Singh, R. and Cattani, C., *Bernoulli collocation method for the third-order Lane-Emden-Fowler boundary value problem*, Appl. Numer. Math. 186 (2023) 100–113.

[27] Singh, R., Guleria, V. and Singh, M., *Haar wavelet quasilinearization method for numerical solution of Emden-Fowler type equations*, Math. Comput. Simul. 174 (2020) 123–133.

[28] Singh, R., Singh, G. and Singh, M., *Numerical algorithm for solution of the system of Emden-Fowler type equations*, Int. J. Appl. Comput. Math. 7(4) (2021) 136.

[29] Singh, R. and Singh, M., *An optimal decomposition method for analytical and numerical solution of third-order Emden-Fowler type equations*, J. Comput. Sci. 63 (2022) 101790.

[30] Turkyilmazoglu, M., *Effective computation of exact and analytic approximate solutions to singular nonlinear equations of Lane-Emden-Fowler type*, Appl. Math. Model. 37(14-15) (2013) 7539–7548.

[31] Yüzbaşı, Ş., *A numerical approach for solving a class of the nonlinear Lane-Emden type equations arising in astrophysics*, Math. Methods Appl. Sci. 34(18) (2011) 2218–2230.

[32] Wazwaz, A.-M., *The variational iteration method for solving systems of equations of Emden-Fowler type*, Int. J. Comput. Math. 88(16) (2011) 3406–3415.

Iranian Journal of Numerical Analysis and Optimization
Vol. 17, No. 1, 2027, pp 25–44
https://doi.org/10.22067/ijnao.2026.99453.1899
https://ijnao.um.ac.ir/

CC BY NC

Research Article

Unraveling the relationship between COVID-19 viral load, disease severity, and innate immune response: Insights from mathematical modeling and sensitivity analysis

M. Lotfi, A. Jabbari*, R. Rahbarghazi and H. Kheiri

Abstract

Coronavirus disease 2019 (COVID-19) created a global health crisis and highlighted the need for a mechanistic understanding of the interactions between SARS-CoV-2 and host immune responses. Although numerous mathematical models have described COVID-19 at the epidemiological scale, fewer studies have examined within-host viral dynamics and innate immune

*Corresponding author

Received 20 June 2026; revised 29 July 2026; accepted 1 August 2026

Mehdi Lotfi
Institute of Bioinformatics, University Medicine Greifswald, Greifswald, Germany.

Azizeh Jabbari
Department of Applied Mathematics, Faculty of Mathematical Sciences, University of Tabriz, Tabriz, Iran. e-mail: a_jabari@tabrizu.ac.ir

Reza Rahbarghazi
Department of Applied Cell Sciences, Faculty of Advanced Medical Sciences, Tabriz University of Medical Sciences, Tabriz, Iran.

Hossein Kheiri
Department of Applied Mathematics, Faculty of Mathematical Sciences, University of Tabriz, Tabriz, Iran.

How to cite this article
Lotfi, M., Jabbari, A., Rahbarghazi, R. and Kheiri, H., Unraveling the relationship between COVID-19 viral load, disease severity, and innate immune response: Insights from mathematical modeling and sensitivity analysis. *Iran. J. Numer. Anal. Optim.*, 2027; 17(1): 25-44. https://doi.org/10.22067/ijnao.2026.99453.1899

mechanisms. In this study, we develop a nonlinear ordinary differential equation model describing interactions among SARS-CoV-2, epithelial cells, resting and activated macrophages, neutrophils, and monocytes. We establish the nonnegativity of solutions, derive the within-host basic reproduction number, and analyze the local and global stability of the COVID-19-free equilibrium. Parameters are calibrated using two clinical datasets representing mild, severe, and intensive care unit (ICU) cases. The calibrated simulations indicate that mild disease is associated with faster viral clearance, more effective immune-mediated viral removal, and better epithelial recovery, whereas severe and ICU cases exhibit delayed clearance, dysregulated neutrophil dynamics, and greater virus-associated immune-cell loss. The corrected normalized sensitivity indices show that viral entry, healthy epithelial-cell renewal, infected-cell lysis, and viral production increase the basic reproduction number, whereas infected-cell recovery, cell loss, and viral clearance reduce it. These findings provide quantitative insight into how innate immune dysregulation may contribute to COVID-19 severity and identify the model mechanisms that most strongly influence within-host viral persistence.

AMS subject classifications (2020): Primary 34A34; Secondary 34D20, 92C37.

Keywords: SARS-CoV-2; COVID-19; innate immune dynamics; within-host modeling; sensitivity analysis.

1 Introduction

Infectious-disease outbreaks have historically posed major threats to human societies. The Plague of Justinian in the sixth century and the Black Death in the fourteenth century caused substantial mortality, while smallpox and cholera have also profoundly affected global health [13, 17]. These examples illustrate the persistent influence of infectious diseases on human survival and social development.

In late 2019, a novel coronavirus, subsequently named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was identified in Wuhan, Hubei Province, China, in association with severe respiratory illness. Despite early containment efforts, the virus spread rapidly and resulted in a global pandemic. By 13 July 2021, more than 186 million confirmed infections and 4 million deaths had been reported worldwide [26]. SARS-CoV-2 primarily affects the respiratory tract and can produce symptoms including fever, cough, fatigue, dyspnea, and myalgia [4].

Coronaviruses are enveloped, single-stranded RNA viruses belonging to the family *Coronaviridae*. They infect several animal species and humans. SARS-CoV-2 is transmitted predominantly through respiratory particles, particularly during close contact. Since the beginning of

the pandemic, extensive research has focused on vaccines, antiviral treatments, and supportive clinical management.

Many epidemiological models have evaluated the effects of self-isolation, mask use, social distancing, hygiene practices, and other public-health interventions on COVID-19 transmission [8, 14, 19, 20, 21]. In contrast, fewer within-host models have investigated viral replication, host–virus interactions, and the effects of immune or therapeutic mechanisms [5, 11, 16]. Epidemiological models support population-level decision-making, whereas within-host models provide complementary insight into cellular and immunological processes that determine viral persistence and disease progression.

The present study develops a within-host mathematical model describing interactions between SARS-CoV-2 and major components of the innate immune response. The model is analyzed with respect to nonnegativity, global existence, the within-host reproduction number, and the stability of the COVID-19-free equilibrium (CFE). We then calibrate unknown parameters using clinical measurements from groups with different disease severities, including mild, severe, and intensive care unit (ICU)-admitted patients. Finally, we compare simulated immune-cell and viral-load trajectories and perform a normalized sensitivity analysis of the reproduction number.

2 Innate immune response

Innate immunity constitutes the first line of defense against invading pathogens. It acts rapidly and nonspecifically to restrict pathogen entry, replication, and dissemination. The innate immune system includes macrophages, eosinophils, natural killer cells, dendritic cells, mast cells, neutrophils, basophils, and other cell populations arising from distinct developmental lineages. Many of these cells originate in the bone marrow, enter the circulation, and migrate to epithelial, mucosal, and connective tissues, where they provide immediate protection.

Granulocytes, including neutrophils, basophils, and eosinophils, contain intracellular granules that are released following activation. Neutrophils and macrophages can also engulf and destroy foreign material through phagocytosis. Enzymatic activity, phagocytosis, cytokine secretion, and cell-to-cell signaling collectively contribute to pathogen clearance and coordinate the early host response. Innate immune cells also interact closely with the adaptive immune system and influence the subsequent development of antigen-specific immunity [23].

In SARS-CoV-2 infection, innate immune mechanisms restrict viral entry, translation, replication, and assembly; promote the recognition and removal of infected cells; and support the initiation of adaptive immune responses [1]. The present model focuses on monocytes, resting and activated macrophages, and neutrophils. These populations were selected because they

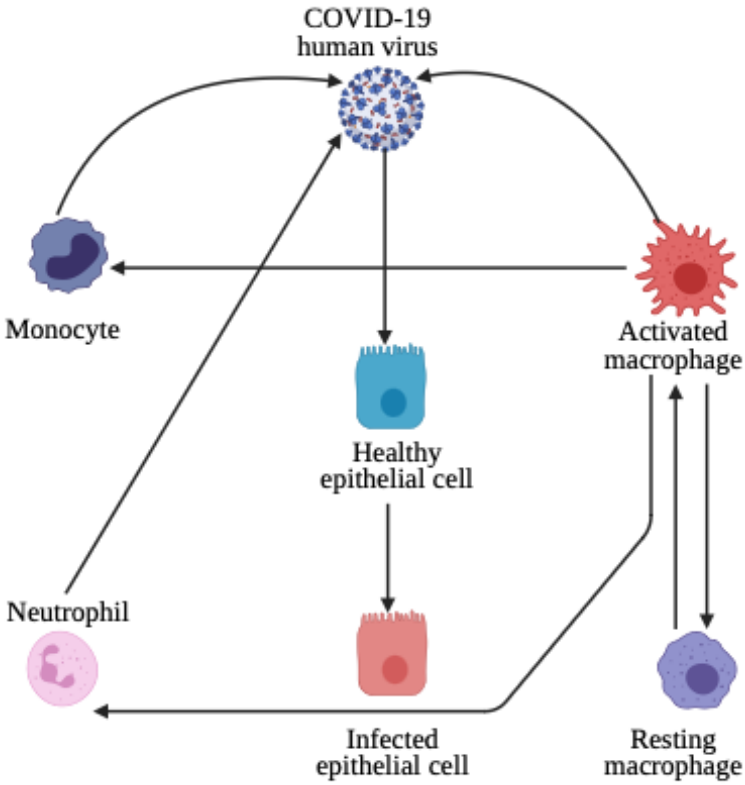


Figure 1: Key immune-cell populations involved in the innate response to SARS-CoV-2. Following exposure to the virus, resting macrophages convert into activated macrophages and can revert to the resting state through macrophage polarization [18]. Activated macrophages release signaling molecules that recruit monocytes from the circulation to the site of infection, thereby increasing macrophage numbers in affected tissues. They also secrete inflammatory mediators that enhance the activity of granulocytes such as neutrophils. The coordinated interactions among these innate immune cells help restrict viral replication and limit the spread of SARS-CoV-2.

are central to the early antiviral response and because clinical data describing neutrophil and monocyte dynamics were available for model calibration. This reduced representation yields a tractable model while retaining key features of the innate response.

3 Model construction

The proposed within-host model describes interactions among SARS-CoV-2, epithelial cells, and selected innate immune-cell populations, as illustrated in Figure 1. The state variables are $S_E(t)$, uninfected epithelial cells; $I_E(t)$, infected epithelial cells; $M_1(t)$, resting macrophages;

$M_2(t)$, activated macrophages; $C_N(t)$, neutrophils; $C_M(t)$, monocytes; and $V(t)$, free viral load. Definitions of the state variables and parameters are provided in Table 1. The within-host dynamics are described by the following nonlinear system of ordinary differential equations:

$$\begin{aligned}
\frac{dS_E}{dt} &= \psi_s - \xi S_E V - \mu_s S_E + \rho I_E,\\
\frac{dI_E}{dt} &= \xi S_E V - \mu_i I_E - \rho I_E,\\
\frac{dM_1}{dt} &= \psi_r - \alpha_r M_1 V - \mu_r M_1 + \kappa M_2,\\
\frac{dM_2}{dt} &= \alpha_r M_1 V - \mu_a M_2 - \kappa M_2 - \nu_a M_2 V,\\
\frac{dC_N}{dt} &= \psi_n M_2 C_N - \mu_n C_N - \nu_n C_N V,\\
\frac{dC_M}{dt} &= \psi_m M_2 C_M - \mu_m C_M - \nu_m C_M V,\\
\frac{dV}{dt} &= \eta \psi_v I_E - \mu_v V - \gamma_c M_2 V - \gamma_n C_N V - \gamma_m C_M V.
\end{aligned} \tag{1}$$

3.1 Biological interpretation of the proposed model

(1) **Healthy epithelial cells:** These cells form the epithelial lining of mucosal surfaces and alveolar structures. They are renewed from progenitor cells at rate ψ_s, become infected through the mass-action term $\xi S_E V$, and are lost naturally at rate $\mu_s S_E$. The term ρI_E represents recovery or replacement of infected epithelial cells before irreversible cytopathic injury.

(2) **Infected epithelial cells:** New infected cells are generated at rate $\xi S_E V$. They are removed by natural or infection-associated loss at rate $\mu_i I_E$ and return to the uninfected compartment through the recovery term ρI_E.

(3) **Resting macrophages:** Resting macrophages are recruited at rate ψ_r and are lost naturally at rate $\mu_r M_1$. Viral exposure activates them at rate $\alpha_r M_1 V$, whereas activated macrophages revert to the resting state at rate κM_2.

(4) **Activated macrophages:** Activated macrophages arise through virus-dependent activation of resting macrophages. Their population declines through natural loss and reversion at rate $(\mu_a + \kappa)M_2$, and through virus-associated loss at rate $\nu_a M_2 V$.

(5) **Neutrophils:** The term $\psi_n M_2 C_N$ represents macrophage-associated neutrophil recruitment or expansion. Neutrophils are lost naturally at rate $\mu_n C_N$ and through virus-associated loss at rate $\nu_n C_N V$.

(6) **Monocytes:** Macrophage-associated monocyte recruitment or expansion is represented by $\psi_m M_2 C_M$. Natural and virus-associated losses occur at rates $\mu_m C_M$ and $\nu_m C_M V$, respectively.

(7) **Free virus:** Viral particles are produced from infected epithelial cells at rate $\eta\psi_v I_E$. They are removed naturally at rate $\mu_v V$ and through the actions of activated macrophages, neutrophils, and monocytes at rates $\gamma_c M_2 V$, $\gamma_n C_N V$, and $\gamma_m C_M V$, respectively.

4 Basic properties of the model

4.1 Existence, positivity, and boundedness

Because the right-hand side of system (1) is locally Lipschitz continuous on $\mathbb{R}^7_+$, a unique local solution exists for every nonnegative initial condition.

Theorem 1. For every nonnegative initial condition, the solution of system (1) remains nonnegative for all times for which it exists. Moreover, the total epithelial-cell population $E = S_E + I_E$, the total macrophage population $Q = M_1 + M_2$, and the viral load V are ultimately bounded. The neutrophil and monocyte variables cannot undergo finite-time blow-up. If $\psi_n \overline{M}_2 < \mu_n$ and $\psi_m \overline{M}_2 < \mu_m$ for an eventual upper bound $\overline{M}_2$ of M_2, then C_N and C_M are also ultimately bounded.

Proof. On the boundary of the nonnegative orthant,

$$\left.\frac{dS_E}{dt}\right|_{S_E=0} = \psi_s + \rho I_E > 0, \qquad \left.\frac{dI_E}{dt}\right|_{I_E=0} = \xi S_E V \geq 0,$$
$$\left.\frac{dM_1}{dt}\right|_{M_1=0} = \psi_r + \kappa M_2 > 0, \qquad \left.\frac{dM_2}{dt}\right|_{M_2=0} = \alpha_r M_1 V \geq 0,$$
$$\left.\frac{dC_N}{dt}\right|_{C_N=0} = 0, \qquad \left.\frac{dC_M}{dt}\right|_{C_M=0} = 0,$$
$$\left.\frac{dV}{dt}\right|_{V=0} = \eta\psi_v I_E \geq 0.$$

Table 1: Definitions of the state variables and parameters used in the proposed model.

Variables	**Description**
$S_E(t)$	Population of uninfected epithelial cells
$I_E(t)$	Population of epithelial cells infected with the virus
$M_1(t)$	Population of resting macrophages
$M_2(t)$	Population of activated macrophages
$C_N(t)$	Population of neutrophils
$C_M(t)$	Population of monocytes
$V(t)$	Viral load (free virus particles)
Parameters	**Description**
ψ_v	Viral output produced from lysed infected cells
η	Rate of lysis of infected epithelial cells
γ_c	Removal of virus particles by macrophages
γ_n	Removal of virus particles by neutrophils
γ_m	Removal of virus particles by monocytes
ψ_s	Renewal of healthy epithelial cells from progenitors
ξ	Rate of viral infection of epithelial cells
μ_s	Turnover (loss) of healthy epithelial cells
μ_i	Turnover (loss) of infected epithelial cells
ψ_r	Recruitment rate of resting macrophages
α_r	Transition of resting macrophages into activated form
μ_r	Loss of resting macrophages
κ	Reversion of activated macrophages to the resting state
μ_a	Loss of activated macrophages
ν_a	Virus-driven loss of activated macrophages
ψ_n	Recruitment of neutrophils induced by macrophages
μ_n	Loss of neutrophils
ν_n	Virus-driven loss of neutrophils
ψ_m	Recruitment of monocytes induced by macrophages
μ_m	Loss of monocytes
ν_m	Virus-driven loss of monocytes
μ_v	Natural clearance rate of free virus

Thus, the vector field is inward-pointing on every boundary hyperplane, and all state variables remain nonnegative.

Let $E = S_E + I_E$ and $\mu_E = \min\{\mu_s, \mu_i\}$. Then

$$\frac{dE}{dt} = \psi_s - \mu_s S_E - \mu_i I_E \leq \psi_s - \mu_E E,$$

which implies

$$\limsup_{t\to\infty} E(t) \leq \frac{\psi_s}{\mu_E}.$$

Similarly, for $Q = M_1 + M_2$ and $\mu_Q = \min\{\mu_r, \mu_a\}$,

$$\frac{dQ}{dt} = \psi_r - \mu_r M_1 - \mu_a M_2 - \nu_a M_2 V \leq \psi_r - \mu_Q Q,$$

and therefore

$$\limsup_{t\to\infty} Q(t) \leq \frac{\psi_r}{\mu_Q}.$$

Since I_E is bounded,

$$\frac{dV}{dt} \leq \eta\psi_v I_E - \mu_v V$$

provides an ultimate bound for V. Because M_2 and V are bounded, the equations for C_N and C_M are scalar linear equations with bounded time-dependent coefficients. Hence, these variables remain finite on every finite time interval, and the local solution extends globally. The additional inequalities stated in the theorem yield exponential comparison bounds and ultimate boundedness of C_N and C_M. □

4.2 Basic reproduction number

The CFE (COVID-19-free equilibrium) is obtained by setting $I_E = M_2 = C_N = C_M = V = 0$ and solving the remaining steady-state equations:

$$E_0 = \left(\frac{\psi_s}{\mu_s}, 0, \frac{\psi_r}{\mu_r}, 0, 0, 0, 0\right).$$

For the infected variables $x = (I_E, V)^T$, the next-generation decomposition at E_0 is

$$\mathcal{F} = \begin{pmatrix} 0 & \xi\psi_s/\mu_s \\ 0 & 0 \end{pmatrix}, \qquad \mathcal{V} = \begin{pmatrix} \mu_i + \rho & 0 \\ -\eta\psi_v & \mu_v \end{pmatrix}.$$

Using the next-generation matrix method [2], the basic reproduction number is

$$R_0 = \rho(\mathcal{F}\mathcal{V}^{-1}) = \frac{\xi\psi_s\eta\psi_v}{(\mu_i + \rho)\mu_s\mu_v}. \tag{2}$$

Biologically, R_0 represents the expected number of newly infected epithelial cells generated by one infected epithelial cell in an otherwise susceptible within-host environment. It combines target-cell availability, viral entry and production, the duration of infected-cell survival, and free-virus persistence.

4.3 Local stability of the CFE

Theorem 2. The CFE E_0 is locally asymptotically stable if $R_0 < 1$ and unstable if $R_0 > 1$.

Proof. At E_0, five eigenvalues of the Jacobian are

$$-\mu_s, \qquad -\mu_r, \qquad -(\mu_a + \kappa), \qquad -\mu_n, \qquad -\mu_m.$$

The remaining two eigenvalues are those of the infected subsystem

$$A = \begin{pmatrix} -(\mu_i + \rho) & \xi\psi_s/\mu_s \\ \eta\psi_v & -\mu_v \end{pmatrix}.$$

The trace of A is negative, and

$$\det(A) = (\mu_i + \rho)\mu_v - \frac{\xi\psi_s\eta\psi_v}{\mu_s} = (\mu_i + \rho)\mu_v(1 - R_0).$$

Consequently, both eigenvalues of A have negative real parts when $R_0 < 1$. When $R_0 > 1$, $\det(A) < 0$, so A has one positive eigenvalue and E_0 is unstable. □

4.4 Global stability of the CFE

Define

$$\Omega_0 = \left\{ (S_E, I_E, M_1, M_2, C_N, C_M, V) \in \mathbb{R}^7_+ : S_E + I_E \leq \frac{\psi_s}{\mu_s} \right\}.$$

If $\mu_i \geq \mu_s$, the set Ω_0 is positively invariant for initial data satisfying $S_E(0) + I_E(0) \leq \psi_s/\mu_s$.

Theorem 3. Within Ω_0, the CFE E_0 is globally asymptotically stable whenever $R_0 \leq 1$.

Proof. Consider the Lyapunov function

$$\mathcal{L} = I_E + \frac{\mu_i + \rho}{\eta\psi_v} V.$$

Along solutions of system (1),

$$\begin{aligned} \dot{\mathcal{L}} &= \xi S_E V - (\mu_i + \rho) I_E \\ &\quad + \frac{\mu_i + \rho}{\eta\psi_v} \left(\eta\psi_v I_E - \mu_v V - \gamma_c M_2 V - \gamma_n C_N V - \gamma_m C_M V \right) \end{aligned}$$

$$\leq \left(\xi S_E - \frac{(\mu_i+\rho)\mu_v}{\eta\psi_v}\right) V.$$

Since $S_E \leq \psi_s/\mu_s$ in Ω_0,

$$\dot{\mathcal{L}} \leq \frac{(\mu_i+\rho)\mu_v}{\eta\psi_v}(R_0-1)V \leq 0$$

whenever $R_0 \leq 1$. By the boundedness results above, each trajectory in Ω_0 is contained in a compact positively invariant subset of the nonnegative state space. The largest invariant subset of $\{\dot{\mathcal{L}} = 0\}$ is characterized by $V = I_E = 0$. The remaining equations then imply

$$S_E \to \frac{\psi_s}{\mu_s}, \qquad M_1 \to \frac{\psi_r}{\mu_r}, \qquad M_2, C_N, C_M \to 0.$$

LaSalle's invariance principle therefore establishes the global asymptotic stability of E_0 in Ω_0. □

5 Simulation and discussion

Numerical simulations were performed in Julia to examine viral-load trajectories, estimate unknown parameters, and evaluate the sensitivity of R_0. Some parameter values were obtained from published studies, as summarized in Table 2, whereas parameters not available from previous within-host COVID-19 models were estimated separately for the two clinical datasets.

Table 2: Fixed parameter values used in the proposed model.

Parameter	Value	Unit	Source
$\eta\psi_v$	$1/19.03$	$\text{virus}\,\text{cell}^{-1}\,\text{day}^{-1}$	[3]
γ_c	1.31	day^{-1}	[3]
ψ_s	4×10^{-3}	$\text{cells}\,\text{mL}^{-1}\,\text{day}^{-1}$	[12]
ξ	5×10^{-9}–5.61×10^{-7}	$\text{mL}\,\text{virus}^{-1}\,\text{day}^{-1}$	[3]
μ_s	0.14	day^{-1}	[22]
μ_i	0.15	day^{-1}	[3]
ψ_r	0.02	day^{-1}	[6]
α_r	0.001	day^{-1}	[6]
μ_r	0.02	day^{-1}	[6]
κ	0.05–0.08	day^{-1}	[6]
μ_a	0.02	day^{-1}	[6]
ν_a	10^{-7}	cell^{-1}	[6]

5.1 Datasets

The first dataset, reported in [24], included 116 confirmed COVID-19 cases in Wenzhou, China, between 19 January and 20 February 2020. Twenty-seven patients were classified as severe, of whom nine required ICU treatment. Severe disease was defined using clinical criteria including elevated respiratory rate, reduced resting oxygen saturation, impaired oxygenation ($PaO_2/FiO_2 \leq 300\,\text{mmHg}$), or rapid radiographic progression. ICU admission was associated with respiratory failure requiring mechanical ventilation, circulatory shock, or multiorgan dysfunction. Neutrophil and monocyte measurements from the severe ICU and severe non-ICU subgroups were used for parameter estimation.

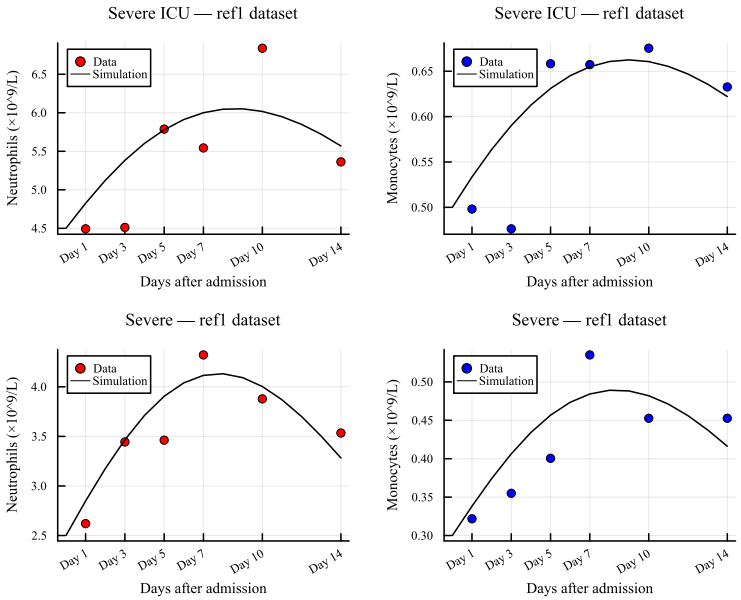
The second dataset, reported in [9], included 61 patients divided into mild (42 patients) and severe (19 patients) groups. Mild cases ranged from limited symptoms without radiographic pneumonia to fever and respiratory symptoms with imaging-confirmed pneumonia. Severe cases met criteria similar to those used in the first dataset. In the remainder of this study, the first and second datasets are denoted by ref1 and ref2, respectively.

5.2 Parameter estimation

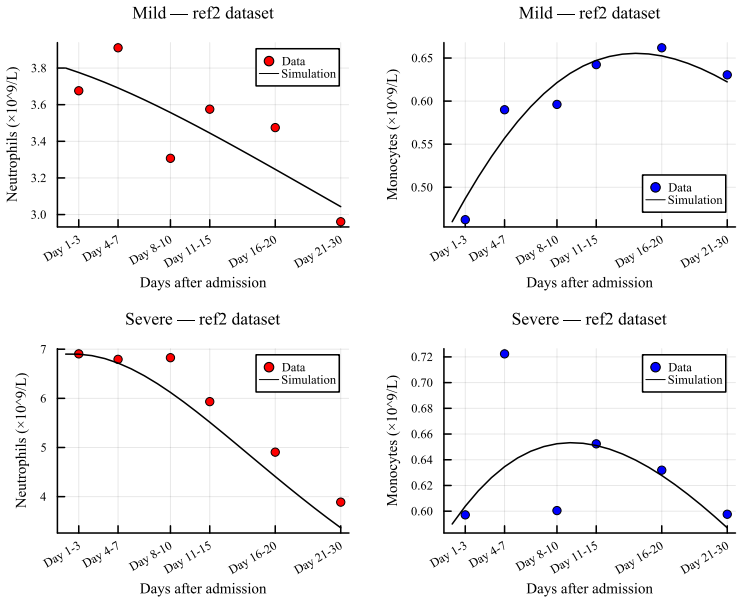
Table 3: Estimated parameter values for the different clinical groups.

Parameter	Severe ICU (ref1)	Severe (ref1)	Mild (ref2)	Severe (ref2)	Unit	Source
γ_n	0.05	0.05	0.1	0.05	day^{-1}	Estimated
γ_m	0.05	0.093	0.1	0.05	day^{-1}	Estimated
ρ	0.05	0.0025	0.05	4.1×10^{-10}	day^{-1}	Estimated
ψ_n	0.17	0.338	0.0096	0.08	model-specific	Estimated
μ_n	0.093	0.196	0.0156	0.076	day^{-1}	Estimated
ν_n	8.5	5.7×10^{-8}	0.5	8.5	model-specific	Estimated
ψ_m	0.15	0.294	0.096	0.053	model-specific	Estimated
μ_m	0.08	0.164	0.0359	0.028	day^{-1}	Estimated
ν_m	0.5	0.5	0.01	0.1	model-specific	Estimated
μ_v	0.03	0.03	0.29	0.03	day^{-1}	Estimated

Unknown parameters were estimated independently for each severity group. For each dataset, the objective function was defined as the sum of squared differences between the observed neutrophil and monocyte measurements and the corresponding model outputs. A genetic algorithm was used to explore the nonlinear parameter space because the objective function may be multimodal and the parameters may interact strongly.



(a)



(b)

Figure 2: Goodness of fit between model simulations and clinical data. (a) Results based on dataset ref1, showing monocyte and neutrophil dynamics in severe and severe ICU patients. (b) Results based on dataset ref2, showing monocyte and neutrophil dynamics in mild and severe patients. In both datasets, the model reproduces the observed temporal patterns, indicating good agreement with the clinical data.

Table 3 summarizes the estimated values, and Figure 2 compares the model simulations with the clinical observations. Because only a limited number of group-level time points were available, the fitted curves should be interpreted as calibrated representations of the reported temporal patterns rather than as individual-patient predictions.

5.3 Estimated parameters and disease severity

Figure 3 highlights differences in the estimated parameters after normalization relative to the mild group. Neutrophil-mediated viral clearance (γ_n) was greatest in the mild ref2 group. Monocyte-mediated viral clearance (γ_m) was also high in the mild ref2 and severe ref1 groups, suggesting that effective monocyte activity may contribute to viral control in some noncritical cases. The epithelial recovery parameter ρ was substantially lower in severe ref1 and nearly zero in severe ref2 patients, indicating impaired epithelial restoration in these groups.

Neutrophil recruitment or expansion (ψ_n) and natural loss (μ_n) were highest in severe ref1 patients, consistent with a highly dynamic and potentially dysregulated neutrophil response. Virus-associated neutrophil loss (ν_n) varied markedly across groups and was high in severe ICU ref1 and severe ref2 patients but negligible in severe ref1. This large between-dataset variability may reflect cohort differences, limited sampling, or practical parameter-identifiability constraints. Monocyte recruitment or expansion (ψ_m) was greater in severe ref1 than in severe ICU ref1 and was higher in mild than severe ref2. Virus-associated monocyte loss (ν_m) was substantially lower in the mild group than in either ref1 severe group. Finally, the natural viral-clearance rate μ_v was much higher in mild ref2 than in the other groups.

Overall, the estimated parameter patterns support a cautious interpretation: mild disease is associated with more efficient viral clearance and epithelial recovery, whereas severe disease is characterized by impaired repair and dysregulated immune-cell turnover. Recruitment estimates do not change monotonically with severity; therefore, elevated recruitment should not automatically be interpreted as an effective response, because excessive innate-cell activation may coexist with tissue injury.

5.4 Initial conditions

The initial-condition vector was ordered as

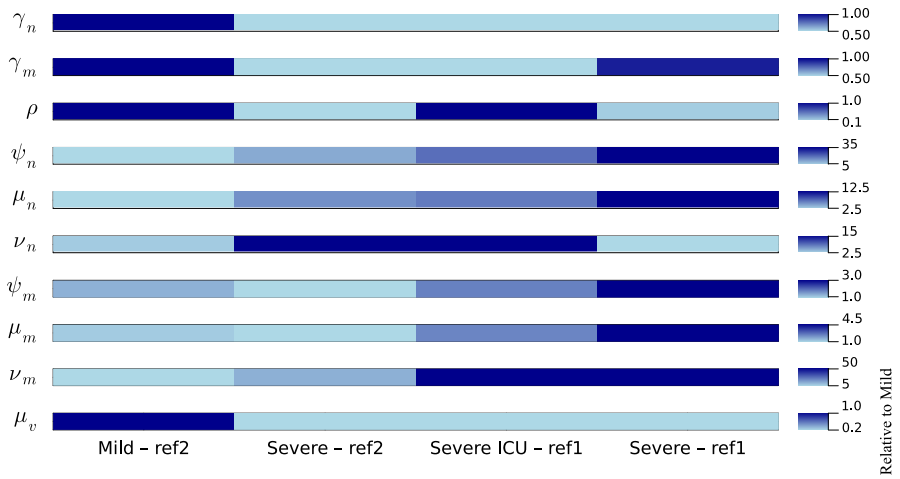


Figure 3: Estimated parameter patterns across the clinical groups. Mild cases show stronger natural viral clearance (μ_v) and more effective epithelial recovery (ρ). Severe and ICU cases show impaired epithelial repair, heterogeneous neutrophil and monocyte recruitment or expansion (ψ_n, ψ_m), and, in several groups, increased virus-associated immune-cell loss (ν_n, ν_m). The recruitment parameters do not vary monotonically with disease severity.

$$u_0 = [S_E(0), I_E(0), M_1(0), M_2(0), C_N(0), C_M(0), V(0)].$$

The values were selected using the reported severity-dependent immune-cell measurements and values adopted in related within-host models [7, 10, 25]. For both severe ICU and severe non-ICU groups in ref1, we used

$$u_0 = [0.4, 3 \times 10^{-8}, 0, 1, 2.5, 0.3, 357 \times 10^{-6}].$$

For the mild ref2 group, we used

$$u_0 = [0.4, 3 \times 10^{-8}, 0, 1, 3.8, 0.46, 357 \times 10^{-6}],$$

whereas the severe ref2 group was initialized with

$$u_0 = [0.4, 3 \times 10^{-8}, 0, 1, 6.9, 0.59, 357 \times 10^{-6}].$$

These initial conditions reproduce the baseline differences reported for the patient groups and should be interpreted as group-level calibration inputs rather than individualized clinical states.

5.5 Dynamics of neutrophils, monocytes, and viral load across severity groups

Figure 4 compares the simulated neutrophil, monocyte, and viral-load trajectories for the four severity groups. Viral load is displayed on a $\log_{10}$ scale. All groups exhibit an early decline, but the mild ref2 group maintains the lowest viral burden and shows the fastest overall clearance. The severe groups display slower clearance and retain higher viral loads over much of the simulated period. Because treatment information was not explicitly included in the model, differences in the simulated decline should be interpreted as consequences of the calibrated parameter sets rather than as direct effects of intensive-care management or specific therapies.

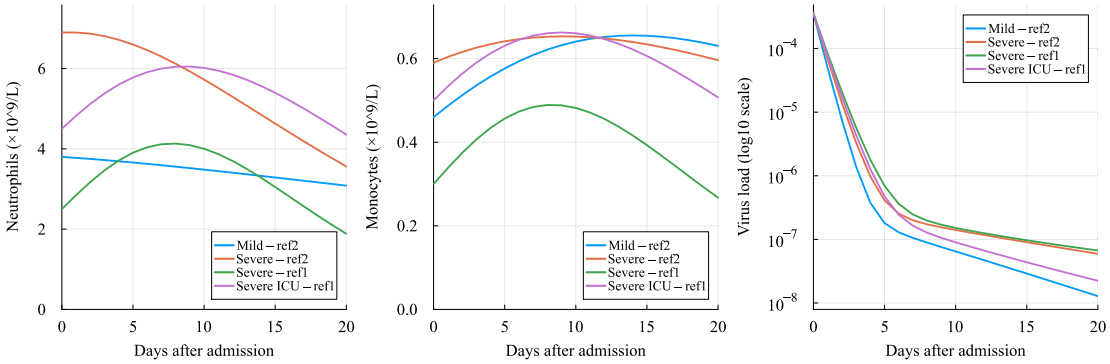


Figure 4: Simulated trajectories of neutrophils (left), monocytes (middle), and viral load (right) for patient groups of varying severity from datasets ref1 and ref2. Mild cases (ref2) show lower and more stable neutrophil counts, sustained monocyte levels, and the fastest viral clearance. Severe and ICU cases exhibit elevated neutrophils, altered monocyte dynamics, and delayed viral clearance, with severe ref1 patients retaining the highest viral load throughout the simulation.

The immune-cell trajectories also differ among groups. Neutrophil levels are lowest and comparatively stable in mild ref2 patients. Severe ref1 patients show a pronounced increase followed by a decline, whereas severe ICU ref1 and severe ref2 patients maintain elevated neutrophil levels. These patterns are consistent with a dysregulated neutrophil response in severe disease, although the simulations alone cannot establish a causal relationship with tissue injury.

Monocyte levels remain comparatively sustained in mild ref2, severe ref2, and severe ICU ref1 groups, whereas severe ref1 patients show lower overall levels. Taken together, the model-calibrated trajectories associate mild disease with rapid viral clearance and more controlled immune-cell dynamics, while severe and ICU disease is associated with delayed clearance and heterogeneous innate immune responses.

5.6 Sensitivity analysis

The normalized forward sensitivity index of R_0 with respect to a parameter ω is defined by

$$\Upsilon_{\omega}^{R_0} = \frac{\partial R_0}{\partial \omega}\frac{\omega}{R_0}.$$

A positive index indicates that an increase in the parameter increases R_0, whereas a negative index indicates that an increase decreases R_0. From equation (2), the indices are

$$\Upsilon_{\xi}^{R_0} = 1, \quad \Upsilon_{\psi_s}^{R_0} = 1, \quad \Upsilon_{\eta}^{R_0} = 1, \quad \Upsilon_{\psi_v}^{R_0} = 1,$$
$$\Upsilon_{\mu_s}^{R_0} = -1, \quad \Upsilon_{\mu_v}^{R_0} = -1, \quad \Upsilon_{\mu_i}^{R_0} = -\frac{\mu_i}{\mu_i+\rho}, \quad \Upsilon_{\rho}^{R_0} = -\frac{\rho}{\mu_i+\rho}.$$

Thus, viral entry (ξ), healthy epithelial-cell renewal (ψ_s), infected-cell lysis (η), and viral output (ψ_v) increase R_0 proportionally. In contrast, greater healthy-cell loss (μ_s), free-virus clearance (μ_v), infected-cell loss (μ_i), and epithelial recovery (ρ) reduce R_0. The indices of ξ, ψ_s, η, ψ_v, μ_s, and μ_v are exactly ± 1 and therefore do not vary among the severity groups. Only the relative negative contributions of μ_i and ρ vary because ρ is group-specific.

For severe ICU ref1 and mild ref2, $\rho = 0.05$ gives

$$\Upsilon_{\mu_i}^{R_0} = -0.75, \qquad \Upsilon_{\rho}^{R_0} = -0.25.$$

For severe ref1, the corresponding indices are approximately -0.984 and -0.016, respectively. For severe ref2, the near-zero value of ρ makes $\Upsilon_{\mu_i}^{R_0}$ approximately -1 and $\Upsilon_{\rho}^{R_0}$ approximately zero. Figure 5 displays these corrected indices.

The sensitivity analysis was performed using the normalized-index approach applied in [15]. The results identify viral entry and production as direct promoters of within-host infection. Conversely, reducing viral entry or production, accelerating free-virus clearance, or improving the recovery and removal of infected epithelial cells is predicted to reduce R_0.

6 Conclusion

We developed and analyzed a within-host mathematical model describing interactions among SARS-CoV-2, epithelial cells, macrophages, neutrophils, and monocytes. The model preserves nonnegativity and admits a biologically meaningful CFE. The within-host basic reproduction number is determined by target-cell availability, viral entry and production, infected-cell recovery or loss, and free-virus clearance. Local stability analysis shows that the CFE is stable when

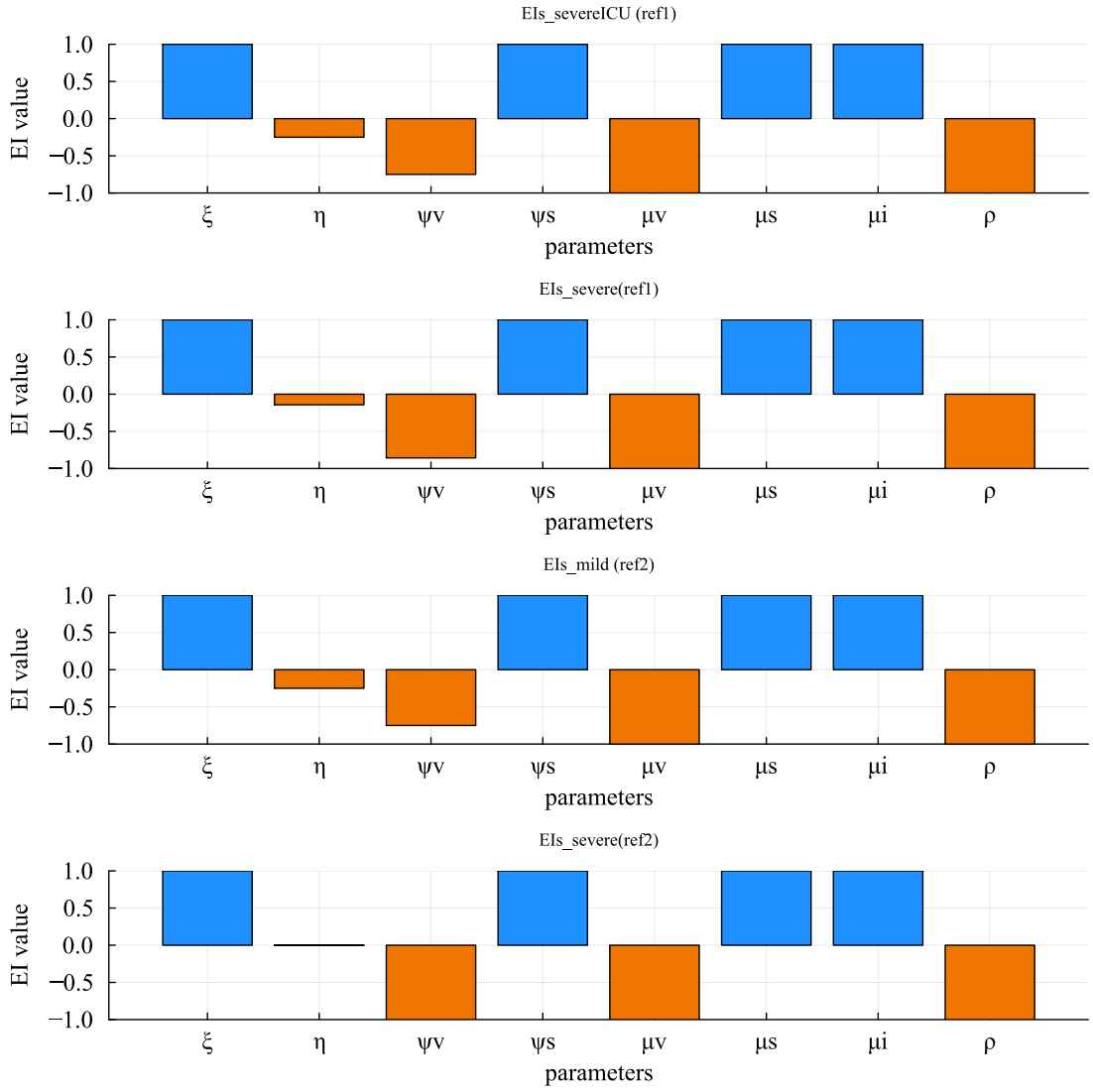


Figure 5: Normalized sensitivity indices of R_0 with respect to the model parameters, represented using a tornado diagram. Positive indices identify parameters that increase within-host viral reproduction, whereas negative indices identify parameters that reduce it.

$R_0 < 1$, and a Lyapunov argument establishes global stability in the stated invariant region when $R_0 \leq 1$.

Calibration to two clinical datasets revealed severity-dependent patterns. Mild disease was associated with faster viral clearance, greater epithelial recovery, and more controlled neutrophil

behavior. Severe and ICU groups exhibited delayed clearance, heterogeneous recruitment, and greater virus-associated immune-cell loss. The corrected normalized sensitivity analysis shows that viral entry and production increase R_0, whereas viral clearance, infected-cell recovery, and cell-loss mechanisms decrease it.

These results demonstrate how within-host mathematical modeling can integrate clinical immune-cell measurements and generate mechanistic hypotheses regarding COVID-19 severity. Extension of the model to adaptive immunity, treatment, host heterogeneity, and patient-level uncertainty may provide broader insight into SARS-CoV-2 and related respiratory viral infections.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Diamond, M.S. and Kanneganti, T.D., *Innate immunity: The first line of defense against SARS-CoV-2*, Nat. Immunol., 23(2) (2022), 165–176. .

[2] Diekmann, O., Heesterbeek, J.A.P. and Metz, J.A.J., *On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations*, J. Math. Biol., 28(4) (1990), 365–382. .

[3] Dogra, P., Ruiz-Ramírez, J., Sinha, K., Butner, J.D., Peláez, M.J., Rawat, M., Yellepeddi, V.K., Pasqualini, R., Arap, W., Sostman, H.D. and Cristini, V., *Innate immunity plays a key role in controlling viral load in COVID-19: Mechanistic insights from a whole-body infection dynamics model*, ACS Pharmacol. Transl. Sci., 4(1) (2021), 248–265. .

[4] Dos Santos, W.G., *Impact of virus genetic variability and host immunity for the success of COVID-19 vaccines*, Biomed. Pharmacother., 136 (2021), 111272. .

[5] Du, S.Q. and Yuan, W., *Mathematical modeling of interaction between innate and adaptive immune responses in COVID-19 and implications for viral pathogenesis*, J. Med. Virol., 92(9) (2020), 1615–1628. .

[6] Eftimie, R. and Hamam, H., *Modelling and investigation of the $CD4^+$ T cells–macrophages paradox in melanoma immunotherapies*, J. Theor. Biol., 420 (2017), 82–104. .

[7] Ghosh, I., *Within-host dynamics of SARS-CoV-2 in humans: Modeling immune responses and antiviral treatments*, SN Comput. Sci., 2(6) (2021), 482. .

[8] Giordano, G., Blanchini, F., Bruno, R., Colaneri, P., Di Filippo, A., Di Matteo, A. and Colaneri, M., *Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy*, Nat. Med., 26(6) (2020), 855–860. .

[9] Guan, J., Wei, X., Qin, S., Liu, X., Jiang, Y., Chen, Y., Chen, Y., Lu, H., Qian, J., Wang, Z. and Lin, X., *Continuous tracking of COVID-19 patients' immune status*, Int. Immunopharmacol., 89(Pt A) (2020), 107034. .

[10] Han, M., Xu, M., Zhang, Y., Liu, Z., Li, S., He, T., Li, J., Gao, Y., Liu, W., Li, T. and Chen, Z., *Assessing SARS-CoV-2 RNA levels and lymphocyte/T-cell counts in COVID-19 patients revealed initial immune status as a major determinant of disease severity*, Med. Microbiol. Immunol., 209 (2020), 657–668. .

[11] Hernandez-Vargas, E.A. and Velasco-Hernandez, J.X., *In-host mathematical modelling of COVID-19 in humans*, Annu. Rev. Control, 50 (2020), 448–456. .

[12] Hernandez-Vargas, E.A. and Velasco-Hernandez, J.X., *In-host modelling of COVID-19 kinetics in humans*, medRxiv preprint (2020). .

[13] Mack, A., Choffnes, E.R., Hamburg, M.A. and Relman, D.A., editors, *Microbial Evolution and Co-Adaptation: A Tribute to the Life and Scientific Legacies of Joshua Lederberg: Workshop Summary*, National Academies Press, Washington, DC, 2009. .

[14] Ivorra, B., Ferrández, M.R., Vela-Pérez, M. and Ramos, A.M., *Mathematical modeling of the spread of coronavirus disease 2019 (COVID-19) taking into account undetected infections: The case of China*, Commun. Nonlinear Sci. Numer. Simul., 88 (2020), 105303. .

[15] Jabbari, A., Lotfi, M., Kheiri, H. and Khajanchi, S., *Mathematical analysis of the dynamics of a fractional-order tuberculosis epidemic in a patchy environment under the influence of re-infection*, Math. Methods Appl. Sci., 46(17) (2023), 17798–17817. .

[16] Li, C., Xu, J., Liu, J. and Zhou, Y., *The within-host viral kinetics of SARS-CoV-2*, Math. Biosci. Eng., 17(4) (2020), 2853–2861. .

[17] McNeill, W.H., *Plagues and Peoples*, Anchor Press, Garden City, NY, 1976.

[18] Murray, P.J., Allen, J.E., Biswas, S.K., Fisher, E.A., Gilroy, D.W., Goerdt, S., Gordon, S., Hamilton, J.A., Ivashkiv, L.B., Lawrence, T. and Locati, M., *Macrophage activation and polarization: Nomenclature and experimental guidelines*, Immunity, 41(1) (2014), 14–20. .

[19] Musa, S.S., Qureshi, S., Zhao, S., Yusuf, A., Mustapha, U.T. and He, D., *Mathematical modeling of COVID-19 epidemic with effect of awareness programs*, Infect. Dis. Model., 6 (2021), 448–460. .

[20] Ndaïrou, F., Area, I., Nieto, J.J. and Torres, D.F.M., *Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan*, Chaos Solitons Fractals, 135 (2020), 109846. .

[21] Peng, L., Yang, W., Zhang, D., Zhuge, C. and Hong, L., *Epidemic analysis of COVID-19 in China by dynamical modeling*, arXiv preprint arXiv:2002.06563 (2020).

[22] Sasmal, S.K., Dong, Y. and Takeuchi, Y., *Mathematical modeling on T-cell-mediated adaptive immunity in primary dengue infections*, J. Theor. Biol., 429 (2017), 229–240. .

[23] Schultze, J.L. and Aschenbrenner, A.C., *COVID-19 and the human innate immune system*, Cell, 184(7) (2021), 1671–1692. .

[24] Sun, S., Cai, X., Wang, H., He, G., Lin, Y., Lu, B., Chen, C., Pan, Y. and Hu, X., *Abnormalities of the peripheral blood system in patients with COVID-19 in Wenzhou, China*, Clin. Chim. Acta, 507 (2020), 174–180. .

[25] Tavares, J.N. and Gomes, E., *COVID-19 immunotherapy: A mathematical model*, Math. Med. Biol., 40(2) (2023), 199–221. .

[26] World Health Organization, *Weekly epidemiological update on COVID-19: 13 July 2021*, World Health Organization, Geneva, 2021. Available from: https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---13-july-2021.

Iranian Journal of Numerical Analysis and Optimization
Vol. 17, No. 1, 2027, pp 45–71
https://doi.org/10.22067/ijnao.2026.98442.1865
https://ijnao.um.ac.ir/

CC BY NC

Research Article

A stable and high-order accurate D-RBF-PU method via hybrid kernels

M. Esmaeilbeigi*, and R. Cavoretto

Abstract

The direct radial basis function partition of unity (D-RBF-PU) method is a powerful mesh-free technique for solving partial differential equations (PDEs). However, its performance is inherently constrained by a critical trade-off between accuracy and numerical stability. Finite-order smoothness kernels, such as thin plate splines and Matérn kernels, ensure well-conditioned systems but severely limit the convergence rate. Conversely, infinitely smooth kernels, such as Gaussians, yield high accuracy but suffer from severe ill-conditioning, leading to computational breakdown at finer discretizations. In this paper, we propose a novel approach using hybrid kernels, strategically combining Gaussian kernels with finitely smooth kernels to effectively bridge this gap.

Numerical experiments on 2D benchmark elliptic PDEs demonstrate that the proposed hybrid formulations achieve an optimal balance. Specifically, hybrid kernels exhibit robust, high-order convergence rates that remain highly competitive with, and in certain structural combinations significantly surpass, the accuracy of pure infinitely smooth bases. Simultaneously, they maintain manageable local and global condition numbers, allowing the method to

*Corresponding author

Received 27 April 2026; revised 30 July 2026; accepted 5 August 2026

Mohsen Esmaeilbeigi
Faculty of Mathematical and Statistical Sciences, Malayer University, Malayer, Iran. e-mail: m.esmaeilbeigi@malayeru.ac.ir

Roberto Cavoretto
Department of Mathematics “Giuseppe Peano”, University of Torino, Torino, Italy. e-mail: roberto.cavoretto@unito.it

How to cite this article
Esmaeilbeigi, M. and Cavoretto, R., A stable and high-order accurate D-RBF-PU method via hybrid kernels. *Iran. J. Numer. Anal. Optim.*, 2027; 17(1): 45-71. https://doi.org/10.22067/ijnao.2026.98442.1865

successfully scale to finer spatial resolutions where traditional infinitely smooth kernels fail. These results confirm that the hybrid kernel approach significantly enhances the robustness and accuracy of the D-RBF-PU method, offering a highly practical solution for large-scale simulations.

AMS subject classifications (2020): Primary 65N35; Secondary 65D05, 41A30, 65F35.

Keywords: Radial basis functions; Partition of unity; Hybrid kernels; Accuracy-stability trade-off.

1 Introduction

Radial basis function (RBF) methods offer a flexible framework for accurate approximation and have been successfully applied in numerous computational settings. These methods have proven particularly valuable in solving partial differential equations (PDEs), which are essential for modeling physical phenomena involving spatial and temporal variations. PDEs describe quantities that fluctuate with respect to geographical and temporal factors, enabling precise encapsulation of system behavior and prediction from initial conditions and external influences [1]. Despite their numerous advantages, global RBF approximations often become inefficient for large-scale problems due to their high computational costs and severe stability challenges.

This limitation has prompted the development of specialized localized techniques, based on the foundational partition of unity method (PUM) pioneered by Babuška and Melenk [16], which led to the development of RBF partition of unity (RBF-PU) methods [20, 7, 2, 3, 13].

In recent years, RBF methods, particularly their localized variants like RBF-PU [15], have transitioned from experimental approaches for small-scale PDE problems to prominent techniques for large-scale PDE simulations. To further enhance the computational efficiency of these localized methods, the direct RBF-PU (D-RBF-PU) method was introduced [17]. This innovative approach utilizes a direct discretization strategy that completely avoids computing the derivatives of PU weights and the lower derivatives of local RBF approximants. By effectively embedding the RBF Finite Difference (RBF-FD) method within a PU framework, the D-RBF-PU method offers a remarkably simpler and more efficient scheme for solving PDEs. Furthermore, it permits the use of discontinuous PU weight functions, which significantly reduces the overall computational expense.

A critical consideration in the D-RBF-PU method is the choice of the basis kernel. In its initial formulation [17], finitely smooth kernels without a shape parameter, such as polyharmonic

splines, were employed to mitigate computational instability. Subsequent advances incorporated finitely smooth kernels with a shape parameter [4]. However, these choices inherently compromise the approximation accuracy. Conversely, utilizing infinitely smooth kernels with a shape parameter (e.g., Gaussian) promises exceptionally high accuracy but introduces severe ill-conditioning and stability breakdowns, especially at finer resolutions.

To circumvent this accuracy-stability dilemma, a rich literature on stabilizing smooth RBF computations has evolved, encompassing advanced methodologies such as the RBF-QR algorithm [12, 11], generalized PDE-solving frameworks [10], the Contour–Padé method [11], and stable basis expansions via Hilbert–Schmidt SVD [8]. While highly powerful, these change-of-basis or complex-plane transformation strategies often induce nontrivial computational overhead or present structural complexities when integrated into highly localized, irregularly bounded partition patches.

In this article, we focus on the D-RBF-PU method for solving benchmark elliptic PDEs, specifically addressing this critical trade-off between accuracy and numerical stability. Departing from the traditional reliance on pure finitely or infinitely smooth kernels, we introduce *hybrid kernels* as a novel and robust solution.

While kernel blending concepts have established antecedents in Gaussian Process regression and machine learning for multi-scale feature representation—including our prior work on hybrid kernels in statistical learning [6]—as well as recent applications in localized meshfree variants [18, 14], their structural mechanism in localized numerical PDE analysis is fundamentally distinct. Specifically, the present work introduces the hybrid formulation not merely as a multi-scale fitting tool, but as a mathematically grounded, localized spectral floor regularizer designed to mitigate severe numerical ill-conditioning in asymmetric PDE collocation systems within the D-RBF-PU architecture. Furthermore, the positioning of this contribution must be explicitly distinguished from previous work in [4]. Although [4] enhanced the flexibility and accuracy of the direct RBF-PU method by transitioning from parameter-free polyharmonic spline kernels to local radial kernels dependent on a shape parameter augmented with local polynomial expansions, it still operated within the framework of pure single-kernel representations. Consequently, it remained susceptible to the classic accuracy-stability trade-off when smooth components are scaled. In contrast, the current study relies on a fundamental structural hybridization of the kernel itself ($\phi_h = \alpha\phi_{\text{inf}} + \beta\phi_{\text{fin}}$). This ensures robust computational resilience even under fixed shape parameters and in the flat limit ($\varepsilon \to 0$), completely bypassing the severe ill-conditioning inherent to pure smooth bases without requiring intricate polynomial augmentations for stability.

These hybrid formulations strategically blend the mathematical strengths of finite-smoothness and infinitely smooth kernels, aiming to achieve an optimal balance that guarantees both high-order accuracy and well-conditioned linear systems.

The remainder of this paper is organized as follows: Section 2 outlines the mathematical formulation of the D-RBF-PU method and discusses the profound impact of kernel selection on its performance. Section 3 introduces the proposed hybrid kernels, detailing how they provide a balanced approach by combining the properties of finite- and infinitely smooth kernels. Section 4 presents comprehensive numerical experiments on 2D benchmark elliptic PDEs, demonstrating significant improvements in both accuracy and stability. Finally, section 5 provides concluding remarks and describes perspectives for future research.

2 The direct RBF-PU method

The D-RBF-PU method is a highly efficient mesh-free local approximation scheme for solving PDEs. This methodology builds upon the standard RBF-PU framework, but introduces a paradigm shift by directly approximating the differential operators. This direct approach substantially simplifies the computational procedures and improves overall efficiency [17]. In this section, we review the mathematical foundations of the D-RBF-PU method, detail its step-by-step construction, and highlight the crucial role of kernel selection in mediating the fundamental trade-off between approximation accuracy and numerical stability.

2.1 Partition of unity framework

Let $\{\Omega_\ell\}_{\ell=1}^{N_c}$ be an open and bounded overlapping covering of a given domain $\Omega \subset \mathbb{R}^d$. A PU subordinate to this covering is a family of compactly supported nonnegative functions $\{w_\ell\}_{\ell=1}^{N_c}$ that satisfy two primary conditions [21]:

1. The support of each weight function is strictly contained within its corresponding patch:

$$\operatorname{supp}(w_\ell) \subseteq \Omega_\ell, \quad \text{for } \ell = 1, \ldots, N_c.$$

2. The sum of the weight functions equals exactly one everywhere in the domain:

$$\sum_{\ell=1}^{N_c} w_\ell(x) = 1, \quad \text{for all } x \in \Omega.$$

Given local approximants s_ℓ of an unknown function u defined in individual patches Ω_ℓ, a globally smooth approximation $s(x)$ is constructed by blending these local functions using the

PU weights:

$$s(x) = \sum_{\ell=1}^{N_c} w_\ell(x) s_\ell(x).$$

In the context of RBF approximations, the local function $s_\ell(x)$ is typically represented as a linear combination of RBFs and polynomial terms:

$$s_\ell(x) = \sum_{j \in J_\ell} c_j \phi(\|x - x_j\|_2) + \sum_{k=1}^{Q} b_k p_k(x),$$

where $J_\ell = \{j : x_j \in X_\ell\}$ represents the indices of the nodes that reside in patch Ω_ℓ, ϕ is the selected radial kernel, and the polynomial terms are necessary to ensure a well-posed system if ϕ is conditionally positive definite (CPD).

2.2 RBF-PU Construction

Let $\phi : \mathbb{R}^d \to \mathbb{R}$ be a CPD RBF of order n. We denote the set of global trial points by $X = \{x_1, \ldots, x_N\} \subset \Omega$, and the localized trial points within a specific patch Ω_ℓ by $X_\ell = X \cap \Omega_\ell$. The local approximation space is formally defined as follows:

$$V_\ell = \text{span}\{\phi(\cdot - x_j) : x_j \in X_\ell\} \oplus \mathbb{P}_{n-1}(\mathbb{R}^d),$$

where $\mathbb{P}_{n-1}(\mathbb{R}^d)$ denotes the space of d-variable polynomials of total degree at most $n-1$, and Q represents its dimension. It is worth noting that if a strictly positive definite (SPD) kernel is employed, the polynomial augmentation can be completely omitted.

Local approximants $s_\ell \in V_\ell$ are determined by interpolating the data values at the local nodes X_ℓ. Enforcing the interpolation conditions, along with the polynomial constraint

$$\sum_{j \in J_\ell} c_j p_k(x_j) = 0,$$

leads to a classic saddle-point linear system:

$$\begin{bmatrix} \Phi & P \\ P^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{c} \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{u}|_{X_\ell} \\ \mathbf{0} \end{bmatrix},$$

where the interpolation matrix entries are given by $\Phi_{ij} = \phi(\|x_i - x_j\|_2)$, and the entries of the polynomial matrix are $P_{jk} = p_k(x_j)$ for a chosen basis $\{p_1, \ldots, p_Q\}$ of $\mathbb{P}_{n-1}$.

2.3 The direct RBF-PU method

Consider a general boundary value PDE problem defined as

$$\begin{aligned} Lu(x) &= f(x), \qquad x \in \Omega, \\ Bu(x) &= g(x), \qquad x \in \Gamma, \end{aligned}$$

where $\Omega \subset \mathbb{R}^d$ is the computational domain, $\Gamma = \partial\Omega$ is its boundary, and L and B are linear differential operators governing the interior and boundary physics, respectively.

In traditional PU methods, evaluating the action of a differential operator L on the global approximation $s(x)$ requires applying the operator to the product $w_\ell(x)s_\ell(x)$. By the product rule, this generates higher-order cross-derivatives of both the PU weights w_ℓ and the local approximants s_ℓ, significantly increasing the mathematical complexity and computational overhead.

The D-RBF-PU method circumvents this bottleneck by proposing a *direct* approximation of the PDE operators. Instead of differentiating the final global function, the method constructs global approximations for the derivatives directly:

$$Lu(x) \approx s^L(x) := \sum_{\ell=1}^{N_c} w_\ell(x) s_\ell^L(x),$$

$$Bu(x) \approx s^B(x) := \sum_{\ell=1}^{N_c} w_\ell(x) s_\ell^B(x).$$

Here, s_ℓ^L and s_ℓ^B are localized RBF approximants of the functions Lu and Bu, respectively. This structure completely bypasses the need to compute any derivatives of the PU weights w_ℓ, thereby significantly reducing the computational cost and simplifying the implementation.

2.4 Accuracy-stability trade-off

A central and persistent challenge in the implementation of the D-RBF-PU method is the inherent trade-off between numerical accuracy and matrix stability, a balance that is dictated entirely by the choice of the base kernel ϕ.

Finite-order smoothness kernels, such as the thin plate spline (TPS), offer robust numerical stability and well-conditioned local matrices. This stability arises largely because they do not rely on a tunable shape parameter. However, their limited algebraic smoothness inherently restricts the maximum achievable convergence rate, making them less suitable for problems demanding extreme high-resolution accuracy [7].

In contrast, infinitely smooth kernels equipped with a shape parameter ε, such as the Gaussian RBF defined by

$$\phi(r) = e^{-\varepsilon^2 r^2},$$

have the potential to achieve spectral or exponential convergence rates under optimal conditions. However, this exceptional accuracy comes at a severe cost driven by two distinct ill-conditioning mechanisms. On one hand, under the flat limit ($\varepsilon \to 0$) at a fixed node distribution, the basis functions physically flatten out, causing the columns of the local interpolation matrices to approach a constant vector and inducing a structural rank deficiency. On the other hand, under the refinement limit ($h \to 0$) at a fixed shape parameter, severe clustering of nodes leads to highly correlated matrix rows due to the dense oversampling of smooth profiles. Both scenarios result in an exponential growth of the condition number, often causing catastrophic numerical breakdowns before the high-order accuracy potential can be practically realized [9].

2.5 Direct approximation and system construction

To compute the local approximants without differentiating the weights, the D-RBF-PU method constructs $s_\ell^L(x)$ and $s_\ell^B(x)$ using generalized Lagrange basis functions:

$$s_\ell^L(x) = \sum_{j \in J_\ell} \mathcal{L}u_j^*(\ell; x)\, u(x_j), \quad \text{and} \quad s_\ell^B(x) = \sum_{j \in J_\ell} \mathcal{B}u_j^*(\ell; x)\, u(x_j).$$

The basis functions $\mathcal{L}u_j^*(\ell; x)$ are obtained by solving a modified local interpolation system in which the target functional is replaced by the action of the operator L on the basis functions. For a given evaluation point $x \in \Omega_\ell$, the system is

$$\begin{bmatrix} \Phi & P \\ P^T & 0 \end{bmatrix} \begin{bmatrix} \mathcal{L}u^*(\ell;x) \\ \mathcal{L}v^*(\ell;x) \end{bmatrix} = \begin{bmatrix} L\phi(\cdot - x)|_{X_\ell} \\ Lp(x) \end{bmatrix}. \tag{1}$$

A strictly analogous linear system is constructed and solved to obtain $\mathcal{B}u^*(\ell;x)$, simply by replacing the differential operator L with the boundary operator B.

The global approximations are then assembled via straightforward summation:

$$s^L(x) = \sum_{\ell=1}^{N_c} \sum_{j \in J_\ell} w_\ell(x)\, \mathcal{L}u_j^*(\ell;x)\, u(x_j),$$

$$s^B(x) = \sum_{\ell=1}^{N_c} \sum_{j \in J_\ell} w_\ell(x)\, \mathcal{B}u_j^*(\ell;x)\, u(x_j).$$

2.6 Final collocation system

Let $Y = Y_\Omega \cup Y_\Gamma$ denote the total set of collocation points, partitioned into interior points Y_Ω and boundary points Y_Γ. The governing PDE and its boundary conditions are enforced discretely at these locations:

$$\begin{aligned} (Lu)(y_k) &= f(y_k), \qquad \text{for } y_k \in Y_\Omega, \\ (Bu)(y_k) &= g(y_k), \qquad \text{for } y_k \in Y_\Gamma. \end{aligned}$$

Substituting the global expansions s^L and s^B into these collocation conditions yields the final global sparse discrete system:

$$\begin{bmatrix} A_L \\ A_B \end{bmatrix} \mathbf{u} = \begin{bmatrix} f|_{Y_\Omega} \\ g|_{Y_\Gamma} \end{bmatrix}, \tag{2}$$

where $\mathbf{u} = [u(x_1), \ldots, u(x_N)]^T$ is the vector of unknown nodal values. The elements of the global matrix are assembled efficiently by accumulating the localized Lagrange weights. Specifically, for an interior point $y_k \in Y_\Omega$:

$$A_L(k,j) = \sum_{\ell \in I(y_k)} w_\ell(y_k)\, \mathcal{L}u_j^*(\ell;y_k),$$

and for a boundary point $y_k \in Y_\Gamma$:

$$A_B(k,j) = \sum_{\ell \in I(y_k)} w_\ell(y_k)\, \mathcal{B}u_j^*(\ell; y_k),$$

where $I(y_k)$ represents the index set of all patches Ω_ℓ that cover the specific evaluation point y_k. Thus

$$\mathbf{A}_{\text{global}}\mathbf{U} = \mathbf{F}. \tag{3}$$

Remark 1. While Proposition 1 in the next section rigorously guarantees the nonsingularity and unisolvence of the localized interpolation systems constructed on each individual partition patch Ω_j, the well-posedness of the final assembled global asymmetric collocation matrix $\mathbf{A}_{\text{global}}$ in (3) warrants careful qualification. In asymmetric Kansa-type RBF collocation methods applied to elliptic operator equations, establishing a universal proof for global unisolvence is a famously difficult open problem. Even for standard smooth kernels, counterexamples involving specialized node configurations where the global matrix becomes singular can theoretically be engineered.

Recently, significant progress has been achieved in this domain; for instance, recent mathematical frameworks (e.g., [19]) have successfully established formal global unisolvence criteria for Kansa-type collocation under elliptic setups specifically using CPD polyharmonic splines (PHS). For the hybrid formulation proposed herein, since the global matrix is a blended sparse assembly of an infinitely smooth component and a PHS-like or Matérn component, we intuitively expect the global system to inherit the robust well-posedness floor of its finitely smooth constituent. However, a formal, mathematically rigorous proof of global unisolvence for hybrid-kernel asymmetric collocation remains a recognized theoretical gap, which we defer to dedicated future investigations in approximation theory.

In summary, the implementation of the D-RBF-PU framework inherently requires the solution of two distinct categories of linear systems:

1. **Local small-scale systems** (1), defined in each overlapping patch Ω_ℓ, to compute operator-dependent Lagrange functions. These dense systems are of size $(|J_\ell| + Q) \times (|J_\ell| + Q)$.

2. **A global collocation system** (2), which is sparse and of size $N \times N$, assembled directly from the local components to produce the final PDE solution.

Crucially, both types of systems are highly susceptible to numerical ill-conditioning, making the strategic mitigation of this issue—such as through the proposed hybrid kernels—vital for the practical viability of the method.

3 Hybrid kernels: A balanced approach

In this section, we propose the integration of *hybrid kernels* as a novel and strategic mechanism to govern the critical trade-off between approximation accuracy and numerical stability in the D-RBF-PU method. By intelligently blending the mathematical strengths of infinitely smooth and finitely smooth RBFs, hybrid kernels offer a highly flexible and robust framework to enhance the performance of mesh-free solvers, particularly for complex elliptic PDEs.

3.1 Formulation of hybrid kernels

A hybrid kernel is fundamentally defined as a positive linear combination of two distinct types of RBFs: One possessing infinite smoothness and the other characterized by finite smoothness. Formally, we construct the hybrid kernel ϕ_h as follows:

$$\phi_h(r) = \alpha\phi_{\text{inf}}(r,\varepsilon) + \beta\phi_{\text{fin}}(r),$$

where

- $\phi_{\text{inf}}(r,\varepsilon)$ is an infinitely smooth RBF equipped with a shape parameter $\varepsilon > 0$.
- $\phi_{\text{fin}}(r)$ is a finitely smooth RBF, which typically does not depend on a shape parameter.
- $\alpha > 0$ and $\beta > 0$ are tunable weight parameters that control the relative contribution of each constituent kernel. Throughout this study, we fix $\alpha = 1$ as a standard normalization choice, since spectral properties and matrix conditioning are strictly governed by the relative weight ratio β/α; altering α merely scales the entire matrix uniformly without affecting its intrinsic condition number.
- $r = \|x - y\|_2$ denotes the standard Euclidean distance between any two points $x, y \in \mathbb{R}^d$.

The core mathematical motivation for this formulation lies in the use of the complementary spectral properties of the two types of kernels. Infinitely smooth kernels provide exponential or spectral convergence rates; however, their associated interpolation matrices exhibit exponentially decaying eigenvalues, leading to rapid ill-conditioning as the grid is refined or as $\varepsilon \to 0$. Conversely, finitely smooth kernels exhibit algebraically decaying eigenvalues, which inherently prevents extreme condition number explosion but limits the maximum achievable accuracy. The hybrid kernel ϕ_h essentially uses ϕ_{fin} to provide a stabilizing "floor" for the matrix spectrum, while allowing ϕ_{inf} to drive the high-order accuracy of the approximation.

3.2 Definiteness, kernel families, and system augmentation

To guarantee the well-posedness and unique solvability of localized collocation systems within the D-RBF-PU framework, it is crucial to establish the definiteness properties of the proposed hybrid kernel ϕ_h. While the sum of two SPD kernels is trivially SPD, the case in which an SPD kernel is combined with a CPD kernel requires careful analysis.

We formalize the characterization of the hybrid formulation by the following proposition:

Proposition 1. Let $\phi_{\text{inf}}(r)$ be an SPD kernel in $\mathbb{R}^d$, and let $\phi_{\text{fin}}(r)$ be a CPD kernel of order n in $\mathbb{R}^d$. Then, for any positive scaling parameters $\alpha > 0$ and $\beta > 0$, the hybrid kernel $\phi_h(r) = \alpha\phi_{\text{inf}}(r) + \beta\phi_{\text{fin}}(r)$ is strictly CPD of order n on $\mathbb{R}^d$.

Proof. Let $X = \{x_1, x_2, \ldots, x_N\} \subset \mathbb{R}^d$ be a set of distinct collocation points. Let A_{inf}, A_{fin}, and A_h denote the evaluation matrices associated with ϕ_{inf}, ϕ_{fin}, and ϕ_h on X, respectively, such that $(A_h)_{ij} = \alpha(A_{\text{inf}})_{ij} + \beta(A_{\text{fin}})_{ij}$.

Let $\Pi_{n-1}(\mathbb{R}^d)$ denote the space of d-variate polynomials of total degree up to $n-1$. Consider any nonzero coefficient vector $c = [c_1, c_2, \ldots, c_N]^T \in \mathbb{R}^N \setminus \{0\}$ that satisfies the polynomial vanishing constraints:

$$\sum_{i=1}^{N} c_i p(x_i) = 0, \quad \forall p \in \Pi_{n-1}(\mathbb{R}^d).$$

We examine the quadratic form associated with the hybrid matrix A_h on this constrained subspace:

$$c^T A_h c = \alpha \left(c^T A_{\text{inf}} c\right) + \beta \left(c^T A_{\text{fin}} c\right).$$

Since ϕ_{inf} is SPD, its interpolation matrix A_{inf} is SPD over the entire Euclidean space, meaning $c^T A_{\text{inf}} c > 0$ holds universally for any nonzero vector c. Moreover, since ϕ_{fin} is CPD of order n, its matrix satisfies $c^T A_{\text{fin}} c \geq 0$ for all nonzero vectors c in $\ker(P^T)$. Given $\alpha > 0$ and $\beta > 0$, the linear combination yields $c^T A_h c > 0$. Consequently, A_h inherits strict conditional positive definiteness of order n, guaranteeing the unisolvence and nonsingularity of the augmented local interpolation systems [7, 21]. □

Remark 2. It is essential to clarify that Proposition 1 strictly guarantees the unisolvence and nonsingularity (invertibility) of the local linear systems. It does not, however, provide a quantitative bound on the spectral condition number. The practical stabilization benefit—namely, capping the growth of condition numbers under spatial refinement—is a distinct physical prop-

erty derived from the heuristic spectral floor analysis on the constrained subspace $V = \ker(P^T)$ presented in Section 3.3.

Based on Proposition 1, the requirement for polynomial augmentation and the resulting size of the local saddle-point system matrix $(|J_\ell| + Q) \times (|J_\ell| + Q)$ on a patch Ω_ℓ are systematically classified as follows:

1. **SPD Constituent Pairing (GA + Matérn C^4):** Since both components are SPD ($n = 0$), the hybrid kernel ϕ_h is unconditionally SPD. The local system is invertible without polynomial augmentation ($Q = 0$), yielding an unaugmented local system matrix of size $|J_\ell| \times |J_\ell|$.

2. **CPD Constituent Pairing of Order $n = 2$ (GA + TPS / GA + Cubic):** Since ϕ_{fin} is CPD of order $n = 2$, ϕ_h is CPD of order 2. Ensuring well-posedness requires augmenting the local basis with degree-1 polynomials ($p \in \Pi_1(\mathbb{R}^d)$). In $d = 2$ spatial dimensions, this introduces $Q = \binom{1+2}{2} = 3$ polynomial terms $(1, x, y)$, expanding the local augmented saddle-point matrix to size $(|J_\ell| + 3) \times (|J_\ell| + 3)$.

To construct this framework under rigorous guidelines, we utilize standard radial families. Commonly employed infinitely smooth kernels include:

$$\begin{aligned} \text{Gaussian (GA):} \quad & \phi_{\text{inf}}(r) = e^{-(\varepsilon r)^2}, \\ \text{Multiquadric (MQ):} \quad & \phi_{\text{inf}}(r) = \sqrt{1 + (\varepsilon r)^2}, \\ \text{Inverse MultiQuadric (IMQ):} \quad & \phi_{\text{inf}}(r) = \frac{1}{\sqrt{1 + (\varepsilon r)^2}}. \end{aligned}$$

In our formulation, the finitely smooth kernels $\phi_{\text{fin}}(r)$ are deliberately kept parameter-free, meaning that $r = \|x - x_j\|_2$ represents the unscaled Euclidean distance without the shape parameter ε. While employing a fixed length scale (or omitting shape parameter tuning) for $\phi_{\text{fin}}(r)$ may be sub-optimal in terms of approximation accuracy compared to an independently scaled constituent, this represents a deliberate stability-oriented trade-off. By decoupling the finitely smooth component from ε, we ensure that $\phi_{\text{fin}}(r)$ provides an invariant, numerically stable spectral baseline that is entirely unaffected by ill-conditioning in the flat limit ($\varepsilon \to 0$) of the infinitely smooth component $\phi_{\text{inf}}(r)$. Furthermore, the Matérn kernel belongs to the global smoothness class $C^4(\mathbb{R}^d)$ for spatial dimensions $d \leq 3$ [7].

3.3 Heuristic error analysis and accuracy-stability trade-off

While a rigorous derivation of native space error bounds for the hybrid formulation within the localized PU framework is deferred to future research, we provide a heuristic argument to illustrate how the blending parameter β systematically governs the trade-off between approximation accuracy and numerical stability.

In standard RBF collocation, theoretical error bounds are governed by the power function $P_{\phi,X}(x)$, whereas numerical stability is dictated by the conditioning of the collocation linear system. To account for both SPD (e.g., Matérn) and CPD (e.g., TPS or Cubic of order m) kernels, the spectral behavior must be evaluated on the polynomial-constrained subspace:

$$V = \ker(P^T) = \left\{c \in \mathbb{R}^N \setminus \{0\} : P^T c = 0\right\},$$

where $P \in \mathbb{R}^{N\times Q}$ represents the local polynomial basis matrix (with $V = \mathbb{R}^N$ for unaugmented SPD kernels). On this subspace V, the constrained quadratic form $c^T Ac$ is strictly positive, matching the augmented saddle-point formulation analyzed in Proposition 1.

We define the constrained minimum eigenvalue (or minimal Rayleigh quotient on V) as follows:

$$\lambda^V_{\min}(A) := \min_{c\in V,\, \|c\|_2=1} c^T Ac > 0.$$

For the infinitely smooth component ϕ_{inf} (e.g., GA), $\lambda^V_{\min}(A_{\text{inf}})$ decays exponentially ($\sim e^{-c/\varepsilon^2}$), leading to severe ill-conditioning and computational collapse as $h \to 0$ or $\varepsilon \to 0$. Conversely, the finitely smooth stabilizer ϕ_{fin} exhibits an algebraic spectral decay, maintaining a strictly positive constrained lower bound $\lambda^V_{\min}(A_{\text{fin}}) \sim \mathcal{O}(h^{2k})$ for a positive integer k determined by the kernel smoothness order.

By construction of the hybrid collocation system $A_h = \alpha A_{\text{inf}} + \beta A_{\text{fin}}$, applying the Rayleigh-Ritz principle restricted to the constrained subspace V yields

$$\lambda^V_{\min}(A_h) \geq \alpha\lambda^V_{\min}(A_{\text{inf}}) + \beta\lambda^V_{\min}(A_{\text{fin}}).$$

When the spatial discretization is highly refined ($h \to 0$), $\lambda^V_{\min}(A_{\text{inf}})$ rapidly drops below the machine precision ($\approx 10^{-16}$). At this critical threshold, the constrained term $\beta\lambda^V_{\min}(A_{\text{fin}})$ becomes dominant, establishing a strict, positive spectral "floor" on V that scales linearly with the blending parameter β. This effectively guarantees the well-posedness of the augmented saddle-

point system solved in Equation (1). Consequently, β functions as a continuous regularization parameter governing the trade-off as follows:

- **Accuracy-Dominated Regime ($\beta \to 0$):** The hybrid kernel behaves almost purely as an infinitely smooth RBF, recapturing high-order convergence rates in the pre-asymptotic regime, while remaining susceptible to numerical saturation on ultra-fine grids.

- **Stability-Dominated Regime (Larger β):** The algebraic stability floor $\beta\lambda^V_{\min}(A_{\text{fin}})$ elevates the constrained spectrum, cushioning the system against spectral collapse. This bounds the condition number of the augmented matrix and enables the D-RBF-PU solver to operate stably on fine spatial resolutions, transitioning the asymptotic convergence rate to the algebraic order of ϕ_{fin}.

This constrained interplay demonstrates how tuning β provides a mathematically sound mechanism to bypass the standard RBF uncertainty principle while preserving full compatibility with Proposition 1.

3.4 Architectural comparison with alternative stabilization techniques

To properly contextualize the proposed hybrid kernel framework within the broader spectrum of RBF stabilization literature, it is instructive to contrast its mathematical and implementation philosophy with established alternatives, such as the RBF-QR algorithm [12], the Contour–Padé method [11], stable bases via Hilbert–Schmidt SVD (HS-SVD) [8], and ε-adaptive strategies [5]. While these methods are highly effective at mitigating the ill-conditioning associated with the flat limit ($\varepsilon \to 0$), they fundamentally differ from the hybrid kernel approach in three major aspects:

1. **Algorithmic Structure and Pre-processing Complexity:** Methods like RBF-QR rely on complex analytical expansions (e.g., spherical harmonics or Chebyshev polynomials) to transition to a stable basis, whereas HS-SVD requires dense matrix decompositions to establish an orthogonal basis in the native space. These operations introduce non-trivial analytical and pre-processing transformations per collocation point. In contrast, the hybrid kernel modifies only the intrinsic entry definitions of the evaluation matrix: $(A_h)_{ij} = \alpha(A_{\text{inf}})_{ij} + \beta(A_{\text{fin}})_{ij}$. Because it operates entirely within the standard RBF

basis, it avoids additional change-of-basis factorizations or special function evaluations during linear system assembly.

2. **Algorithmic Simplicity and Implementation Ease:** Implementing change-of-basis techniques (RBF-QR/HS-SVD) or complex-plane contour integrations (Contour–Padé) requires a substantial rewrite of the standard RBF interpolation pipeline. This restricts their immediate use as "plug-and-play" components in pre-existing industrial PDE codes. Conversely, the hybrid kernel formulation requires no modifications to the overarching linear solver or the localized matrix assembly routines, offering an out-of-the-box, structurally straightforward stabilization tool for practitioners.

3. **Compatibility with Localized Formulations (D-RBF-PU):** Most classical stabilization techniques were originally formulated for global RBF interpolation or specialized geometries (e.g., spheres or torus domains). Extending RBF-QR or Contour–Padé into highly localized, irregularly bounded patches within a PU framework introduces severe mathematical and algorithmic hurdles. The hybrid kernel framework bypasses this geometric restriction entirely, as its stabilizing spectral floor acts symmetrically and isotropically across any local node configuration, regardless of the boundary geometry.

3.5 Integration into the D-RBF-PU Framework

The D-RBF-PU method faces a persistent operational challenge: traditional finitely smooth kernels ensure stable local matrices but yield suboptimal accuracy, while infinitely smooth kernels yield superior accuracy but suffer from computational breakdown at fine resolutions. By adopting ϕ_h, practitioners can actively navigate this trade-off. Increasing the weight α emphasizes accuracy by leveraging the rapid convergence of ϕ_{inf}, whereas increasing β injects robustness and limits the growth of the condition number through ϕ_{fin}.

To integrate hybrid kernels into the D-RBF-PU architecture, we seamlessly replace the standard RBF ϕ with the hybrid formulation ϕ_h. In each local patch Ω_ℓ, the basic approximant is redefined as:

$$s_\ell(x) = \sum_{j \in J_\ell} c_j \phi_h(\|x - x_j\|_2) + \sum_{k=1}^{Q} b_k p_k(x).$$

For the direct approximation of differential operators, local expansions take the form $s_\ell^L(x) = \sum_{j \in J_\ell} \mathcal{L}u_j^*(\ell; x)u(x_j)$ and $s_\ell^B(x) = \sum_{j \in J_\ell} \mathcal{B}u_j^*(\ell; x)u(x_j)$. The modified generalized Lagrange

functions are computed by solving the localized system:

$$\begin{bmatrix} \Phi_h & P \\ P^T & 0 \end{bmatrix} \begin{bmatrix} \mathcal{L}u^*(\ell;x) \\ \mathcal{L}v^*(\ell;x) \end{bmatrix} = \begin{bmatrix} L\phi_h(\cdot - x)|_{X_\ell} \\ Lp(x) \end{bmatrix}.$$

A significant computational advantage of this approach is that, due to the linearity of the differential operators L and B, the action on the hybrid kernel is computed trivially by superposition:

$$L\phi_h(r) = \alpha L\phi_{\text{inf}}(r,\varepsilon) + \beta L\phi_{\text{fin}}(r).$$

Therefore, evaluating the vectors on the right-hand side does not require complex derivations beyond those already established for standard kernels.

Finally, the globally approximated fields remain consistent with the original formulation:

$$s^L(x) = \sum_{\ell=1}^{N_c} w_\ell(x)\, s_\ell^L(x),$$
$$s^B(x) = \sum_{\ell=1}^{N_c} w_\ell(x)\, s_\ell^B(x),$$

and the global sparse collocation system is assembled precisely as detailed in Section 2. By transitioning to the operational matrices A_L and A_B derived from ϕ_h, the overall computational pipeline benefits from significantly improved conditioning at the local level, directly translating to enhanced stability and solvability of the final global system.

4 Numerical experiments and discussion

In this section, we present a comprehensive set of numerical experiments designed to systematically evaluate the performance of the proposed D-RBF-PU method. The primary objective of these experiments is to empirically demonstrate the inherent accuracy-stability trade-off associated with classical RBFs and to validate the efficacy of the proposed hybrid kernel approach in overcoming these limitations.

To this end, we consider a benchmark linear elliptic PDE in a two-dimensional domain $\Omega = [0,1]^2$, governed by

$$Lu = -\Delta u + u = f \quad \text{in } \Omega.$$

Mixed boundary conditions are imposed to rigorously test the method: Dirichlet conditions are prescribed along the vertical boundaries ($x = 0$ and $x = 1$), while Neumann conditions are applied on the horizontal edges ($y = 0$ and $y = 1$). The forcing term f and the boundary data are analytically extracted from the exact solutions chosen. We investigate two distinct test cases:

- **Test case 1 (The Franke Function):** This widely used benchmark is characterized by severe local variations, steep gradients, and multiple peaks, making it a challenging test for mesh-free interpolants. The exact solution is given by

$$\begin{aligned} u(x,y) = &\frac{3}{4}\exp\left(-\frac{1}{4}\left((9x-2)^2+(9y-2)^2\right)\right) \\ &+\frac{3}{4}\exp\left(-\frac{1}{49}(9x+1)^2-\frac{1}{10}(9y+1)^2\right) \\ &+\frac{1}{2}\exp\left(-\frac{1}{4}\left((9x-7)^2+(9y-3)^2\right)\right) \\ &-\frac{1}{5}\exp\left(-\left((9x-4)^2+(9y-7)^2\right)\right). \end{aligned}$$

- **Test case 2 (A Smooth Trigonometric Function):** To evaluate the spectral convergence properties on a globally smooth profile, we employ

$$u(x,y) = \sin(\pi x)\cos\left(\frac{\pi y}{2}\right).$$

Spatial discretization utilizes uniformly distributed collocation points with a nominal node spacing parameter $h = (c_d N)^{-1/d}$, which serves as a global refinement indicator for convergence analysis and plotting purposes. While the rigorous geometric fill distance is classically defined as $h_{X,\Omega} = \sup_{x\in\Omega}\min_{x_j\in X}\|x - x_j\|_2$, for our uniformly distributed trial points, the parameter h provides a computationally straightforward and asymptotically equivalent measure of the discretization scale. Here, N represents the total number of trial points, $d = 2$ is the spatial dimension, and $c_d = \mathrm{vol}(B)/\mathrm{vol}(\Omega)$, where B denotes the bounding box of Ω. For the standard test domain $\Omega = [0,1]^2$, the bounding box coincides with the domain itself ($B = \Omega$), yielding $c_d = 1$ and an effective spatial fill distance of $h = N^{-1/2} = 1/\sqrt{N}$. This explicit relation maps the nodal spacing h directly to the refinement parameter $\sqrt{N}$ depicted on the horizontal axes of our convergence figures. We progressively refine the grid using $N \in \{121, 225, 441, 841, 1681\}$, which corresponds to $\sqrt{N} \in \{11, 15, 21, 29, 41\}$ (or spatial resolution $h \in [1/11, 1/41]$).

To quantitatively assess performance, we monitor three key metrics:

1. The relative ∞-norm error: $E_\infty = \|\mathbf{s} - \mathbf{u}\|_\infty / \|\mathbf{u}\|_\infty$, where $\mathbf{s}$ and $\mathbf{u}$ are the numerical and exact solution vectors, respectively.

2. The numerical convergence order, derived from the slope of the linear regression in the log-log plot of the error versus h.

3. The condition numbers. In the subsequent tables, "Local Cond." denotes the condition number of the local interpolation matrices (averaged across all PU patches), while "Global Cond." reflects the condition number of the final assembled sparse collocation matrix.

To quantify the asymptotic accuracy of the numerical schemes, the global convergence orders p reported in the performance profiles are systematically estimated via a linear least-squares curve fitting applied to the log-log error data against the grid refinement indicator $\sqrt{N}$. Formally, the convergence rate is determined as $p = -a_1$, where a_1 is the optimal slope satisfying the first-degree polynomial relationship $\log(E_\infty) = a_1 \log(\sqrt{N}) + a_2$. For the stable finite-smoothness and proposed hybrid kernel frameworks, this linear regression spans the complete sequence of grid refinements up to $N = 1681$. Conversely, for the pure infinitely smooth configurations, the regression interval is truncated to the stable, pre-breakdown range ($N \leq 441$) to prevent numerical saturation from biasing the estimated convergence rate. All simulations were executed in MATLAB on a workstation equipped with an Intel Core i7 CPU (4.00 GHz) and 16 GB of RAM.

4.1 Baseline: Finitely smooth kernels

Our first set of experiments investigates finitely smooth kernels (Matérn, TPS, and Cubic). Theoretical expectations suggest that these kernels should provide robust numerical stability at the cost of algebraic (lower-order) convergence rates.

Table 1: Performance profile of finitely smooth kernels for Test Case 1 in 2D space ($d = 2$). Matrix size indicates the dimension $(|J_\ell| + Q) \times (|J_\ell| + Q)$ of local collocation systems.

Kernel	Poly. Deg.	N range	Conv.	Local Cond.	Global Cond.
Matérn (C^4)	None ($n=0$)	121–1681	2.27	5e8–5e11	5e4–3e6
TPS	Deg. 1 ($n=2$)	121–1681	1.32	2e6–1e9	8e4–5e6
Cubic	Deg. 1 ($n=2$)	121–1681	2.25	6e7–6e10	2e4–3e6

As illustrated in Table 1 for Test case 1, the numerical results align perfectly with theoretical predictions. The Matérn and Cubic kernels exhibit moderate convergence orders of approximately 2.27 and 2.25, respectively, while the TPS kernel yields a slower rate of 1.32. Notably, the method successfully computes solutions up to the finest grid ($N = 1681$). Although the local

condition numbers increase with grid refinement, they remain bounded within a manageable algebraic limit (up to $\sim 10^{11}$). Furthermore, the PU blending ensures that the global sparse system remains exceptionally well-conditioned (peaking at 10^6), allowing for efficient iterative or direct solutions.

4.2 The stability bottleneck: Infinitely smooth kernels

Next, we evaluate infinitely smooth kernels (GA, MQ, and IMQ), which are historically favored for their spectral convergence but are notorious for severe ill-conditioning as $h \to 0$.

Table 2: Performance profile of pure infinitely smooth kernels. Note: Empirical convergence orders are evaluated over the pre-breakdown regression window $N \in [121, 441]$, as finer grids suffer from severe numerical ill-conditioning.

Kernel	N range	Conv. Order	Local Cond.	Global Cond.
GA ($\varepsilon=2$)	121–441	4.38	2e12–3e16	1e4–9e5
MQ ($\varepsilon=2$)	121–441	3.95	2e8–3e12	4e4–5e5
IMQ ($\varepsilon=2$)	121–441	3.95	3e7–1e11	2e4–4e5

Table 2 reveals the stark contrast between accuracy and stability. For Test case 1, these kernels deliver impressive convergence orders (4.38 for GA, 3.95 for MQ/IMQ). However, this rapid convergence is accompanied by an exponential deterioration in the local condition numbers. For the GA kernel at merely $N = 441$, the local condition number hits the machine precision limit ($\sim 10^{16}$). For any $N > 441$, the local interpolation matrices become numerically singular, preventing the extraction of the generalized Lagrange functions and causing the entire D-RBF-PU algorithm to fail. This completely prohibits the use of pure infinitely smooth kernels on fine resolutions.

4.3 Breaking the barrier: The hybrid kernel approach

To resolve the impasse demonstrated in the previous subsections, we deploy the proposed hybrid kernels, combining the highly accurate GA kernel with stabilizing finitely smooth kernels. The weighting parameters α and β, together with the shape parameter ε, are optimally tuned to restrict the condition number growth while preserving the influence of the GA kernel.

Table 3 summarizes the main numerical results obtained in this study. To ensure a fair and transparent comparison, we note that the reported empirical convergence orders are computed

Table 3: Performance profile of hybrid kernels for Test Case 1.

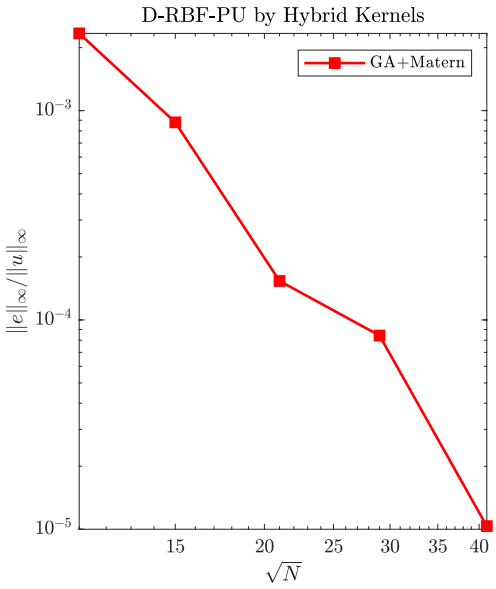
Kernel Formulation	N range	Conv.	Local Cond.	Global Cond.
GA + Matérn ($\varepsilon=5$, $\alpha=1$, $\beta=10^{-8}$)	121–1681	3.68	1e11–5e15	3e4–2e6
GA + TPS ($\varepsilon=2.5$, $\alpha=1$, $\beta=10^{-6}$)	121–1681	5.10	3e10–1e15	2e4–2e6
GA + Cubic ($\varepsilon=2.5$, $\alpha=1$, $\beta=10^{-8}$)	121–1681	4.24	2e10–3e15	2e4–1e6

Note: Empirical convergence orders are evaluated over the full grid refinement window $N \in [121, 1681]$, made possible by the structural stability of the hybrid spectral floor.

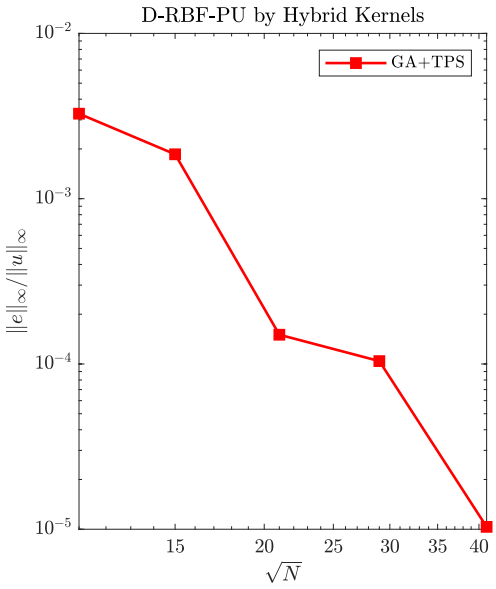
via linear regression over the stable regime of each configuration: for pure smooth kernels, the regression window is truncated to $N \leq 441$ due to catastrophic ill-conditioning at finer resolutions, whereas for hybrid formulations, the window encompasses the full resolution range up to $N = 1681$. By introducing a small fractional weight (β) of a finite-smoothness kernel, we successfully extend the solvability of the method up to the finest grid ($N = 1681$), a feat impossible with the pure GA kernel. Remarkably, this structural stability does not degrade accuracy; instead, convergence orders remain exceptionally high. The GA+TPS combination achieves an outstanding convergence order of 5.10, outperforming both of its individual constituents. Local condition numbers are strictly capped below the threshold of total numerical singularity (up to 10^{15}), and global sparsity maintains the overarching system highly tractable.

To further substantiate these findings across different mathematical landscapes, we analyzed the hybrid kernels using Test case 2. Figure 1 visually maps the convergence trajectories, showing relative ∞-norm error versus $\sqrt{N}$.

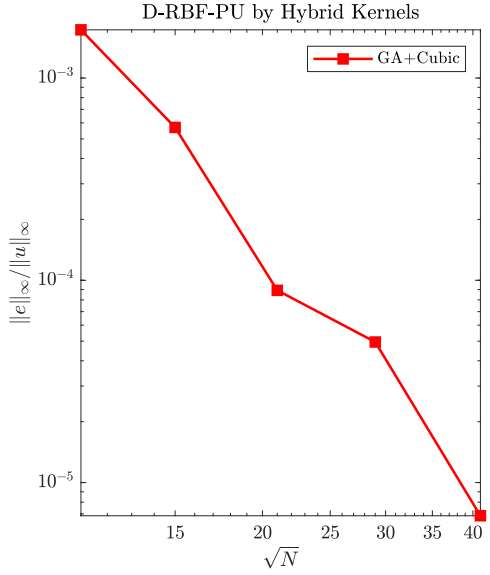
The subplots in Figure 1 uniformly confirm that the hybrid mechanism is robust and independent of the specific underlying exact solution. Convergence orders consistently exceed 4.0, verifying that the hybrid kernels seamlessly blend the spectral decay of the GA component with the numerical resilience of the finite-smoothness component. Note that the local and global condition numbers for Test case 2 are omitted here for brevity, as they follow a nearly identical growth trajectory and remain within the same orders of magnitude as those reported for Test case 1 in Table 3 due to the identical spatial node configurations. Ultimately, these results provide strong empirical evidence that the hybrid D-RBF-PU framework is a highly practical, stable, and accurate tool for large-scale PDE simulations.



(a) GA + Matérn ($\varepsilon = 2$, $\alpha = 1$, $\beta = 10^{-6}$)
Convergence Order = 4.01



(b) GA + TPS ($\varepsilon = 2$, $\alpha = 1$, $\beta = 10^{-8}$)
Convergence Order = 4.38



(c) GA + Cubic ($\varepsilon = 1.6$, $\alpha = 1$, $\beta = 10^{-6}$)
Convergence Order = 4.12

Figure 1: Convergence behavior of hybrid kernels for Test Case 2. The plots track the decay of the relative ∞-norm error ($\|e_h\|_\infty/\|u\|_\infty$) against grid refinement ($\sqrt{N}$). All hybrid combinations reliably maintain high-order convergence without succumbing to ill-conditioning. Note that the local and global condition numbers for Test Case 2 are omitted here for brevity, as they follow a nearly identical growth trajectory and remain within the same orders of magnitude as those reported for Test Case 1 in Table 3 due to the identical spatial node configurations.

4.4 Discussion on convergence and implementation aspects

To better appreciate the significance of the results presented in Tables 1–3, it is instructive to contrast the characteristics of the proposed hybrid kernel framework with existing literature in the domain of RBF-PUMs and D-RBF-PU formulations. Traditional RBF-PUM applications utilizing pure finitely smooth profiles, such as the Matérn C^4 or Cubic families, exhibit highly predictable but strictly limited algebraic convergence rates, seldom exceeding order 3.0 in practical spatial discretizations [3]. On the other hand, achieving high-order convergence rates exceeding 4.0 typically mandates the deployment of infinitely smooth kernels like GAs or MQs. However, as demonstrated in Table 2 and widely documented in classical mesh-free theory [7], these pure infinitely smooth configurations inevitably succumb to catastrophic ill-conditioning in the flat limit or under progressive grid refinement.

The hybrid kernel strategy introduced in this study effectively bridges this long-standing gap. By embedding a parameter-free finite-smoothness baseline, the formulation yields convergence rates as high as 5.10 (Table 3) while successfully maintaining complete numerical stability up to the finest grid ($N = 1681$). From an implementation perspective, this framework offers a distinct structural advantage over specialized stabilization paradigms such as the RBF-QR or RBF-GA algorithms. While those techniques successfully bypass ill-conditioning, they require nontrivial modifications to compute expanded basis transformations, particularly when extending to multi-dimensional boundary value problems. Because the proposed hybrid framework keeps matrix condition numbers within manageable bounds via a structural stability floor, it operates seamlessly with standard, unaugmented linear system solvers. This eliminates the need for change-of-basis algorithms, rendering the proposed scheme both highly accurate and remarkably straightforward to implement compared to standard stabilized RBF-PUM variants.

4.5 Parameter sensitivity and practical guidance

To address the crucial question of parameter selection and to ensure full reproducibility of the reported results, we first clarify the procedure used to select the shape parameter ε and blending weight β (fixing $\alpha = 1$ as a standard system normalization). To ensure a fair comparison based on the best achievable performance of each architectural paradigm, each method (both pure baseline kernels and hybrid configurations) was evaluated using its individually optimized shape parameter.

Specifically, for each kernel combination and test problem, the reported pairs (ε, β)—or ε alone for pure baselines—were systematically identified via an exhaustive grid search over $\varepsilon \in [1, 10]$ and $\beta \in \{10^{-4}, 10^{-5}, \ldots, 10^{-10}\}$, selecting the combination that minimized the Root Mean Square (RMS) error.

It is important to emphasize that this comparative strategy evaluates each formulation under its optimal operational regime. For the pure GA baseline, reducing the shape parameter on fine spatial discretizations causes catastrophic numerical breakdown due to severe ill-conditioning, forcing its optimal operational value to settle around $\varepsilon = 2$. Conversely, the spectral floor provided by the hybrid formulation allows the solver to operate stably at smaller shape parameters without matrix singularity. Consequently, the performance reported in Tables 2–3 reflects a comparison between separately optimized configurations, highlighting how hybrid stabilization enables practitioners to safely access smaller shape parameter regimes that remain computationally intractable for pure infinitely smooth RBFs.

This empirical analysis reveals that the D-RBF-PU method is highly robust and operates successfully over a remarkably wide parameter plateau rather than requiring a narrow, hypersensitive window. Because optimal parameters are selected independently for each kernel formulation and boundary value configuration, the interplay between ε and β is governed by three distinct operational regimes:

1. **The Under-Stabilized Regime ($\beta < 10^{-11}$):** If the blending parameter is chosen too small, the finite-smoothness kernel fails to inject a sufficient spectral floor. Consequently, the exponential decay of the GA component dominates, leading to severe ill-conditioning and early matrix singularity on fine resolutions ($N \geq 841$).

2. **The Optimal Blending Plateau ($10^{-9} \leq \beta \leq 10^{-5}$):** Within this wide range, the hybrid kernel achieves an ideal equilibrium. The local condition numbers are bounded safely below the computational breakdown threshold ($\sim 10^{15}$), while the high-order convergence properties driven by ϕ_{inf} are fully preserved. Across different test problems and kernel pairings (e.g., GA+Matérn, GA+TPS, GA+Cubic), the independently optimized blending weights consistently fall within $10^{-8} \leq \beta \leq 10^{-6}$. Minor shifts in optimal values are dictated by problem geometry and kernel smoothness: sharp, multi-peaked profiles favor slightly smaller β values to preserve maximum spectral accuracy, whereas globally smoother solutions accommodate larger stabilization floors.

3. **The Saturation Regime ($\beta > 10^{-3}$):** When β becomes overly dominant, the algebraic convergence behavior of the finitely smooth kernel overrides the high-order accuracy of the infinitely smooth component. In this case, the method remains perfectly stable, but

the convergence order drops asymptotically toward the baseline algebraic rates reported in Table 1.

For practical engineering implementations where the exact analytical solution is unavailable, we recommend a simple two-step heuristic workflow: first, choose a moderate shape parameter based on the local node density, and second, initialize the blending parameter at a safe baseline of $\beta = 10^{-7}$. For fully automated, data-driven applications, practitioners can seamlessly integrate a localized variant of Rippa's algorithm or the Leave-One-Out Cross-Validation (LOOCV) framework [7], minimizing the local residual norm across individual partition patches to systematically extract optimal (ε, β) pairs without manual intervention.

5 Conclusion

In this paper, we investigate the integration of hybrid kernels within the D-RBF-PU method to resolve the fundamental accuracy-stability trade-off in solving elliptic PDEs. By linearly combining infinitely smooth and finitely smooth kernels, the proposed approach successfully leverages the complementary properties of both classes: it achieves the high convergence rates characteristic of infinitely smooth kernels while utilizing the localized nature of finitely smooth kernels to keep the condition numbers of local interpolation matrices at strictly manageable levels.

Numerical experiments in benchmark test cases clearly demonstrated the distinct advantages of this hybrid formulation. While pure infinitely smooth kernels suffered from severe ill-conditioning and computational breakdown during grid refinement, the hybrid combinations reliably maintained high-order convergence without succumbing to numerical instability. These results confirm that hybrid kernels significantly enhance the robustness of the D-RBF-PU method, making it a practical and highly efficient tool for large-scale PDE simulations where both extreme accuracy and stability are required.

Beyond these foundational results, this study laid the groundwork for a more flexible and intelligent computational framework. Future work will focus on developing an adaptive hybrid strategy designed to dynamically employ different combinations of kernels and optimize the shape parameter (ε) alongside the weight parameters (α and β) based on local nodal distributions, problem dimensionality, or specific computational constraints. Furthermore, extending this methodology to other classes of PDEs (such as parabolic or hyperbolic equations) and conducting a rigorous theoretical analysis of the underlying stability and convergence bounds remain critical directions for future exploration.

Acknowledgements

The authors would like to express their sincere gratitude to the handling editor and the anonymous reviewers for their constructive comments and valuable suggestions, which have significantly improved the quality and presentation of this manuscript.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Bleecker, D. and Csordas, G. *Basic partial differential equations*, Chapman and Hall/CRC, 1995.

[2] Cavoretto, R. *A numerical algorithm for multidimensional modeling of scattered data points,* Comput. Appl. Math. 34 (2015), 65–80.

[3] Cavoretto, R., De Marchi, S., De Rossi, A., Perracchione, E. and Santin, G. *Partition of unity interpolation using stable kernel-based techniques,* Appl. Numer. Math. 116 (2017), 95–107.

[4] Cavoretto, R., De Rossi, A. and Haider, A. *A shape-parameterized RBF-partition of unity technique for PDEs,* Appl. Math. Lett. 163 (2025), 109453.

[5] Cavoretto, R., De Rossi, A. and Lancellotti, S. *Bayesian approach for radial kernel parameter tuning,* J. Comput. Appl. Math. 441 (2024), 115716.

[6] Esmaeilbeigi, M. and Cheraghi, M. *Hybrid kernel approach to improving the numerical stability of machine learning for parametric equations with Gaussian processes in the noisy and noise-free data assumptions,* Eng. Comput. 40 (2024), 761–794.

[7] Fasshauer, G.E. *Meshfree approximation methods with MATLAB*, Interdisciplinary Mathematical Sciences, Vol. 6, World Scientific Publishing Co., 2007.

[8] Fasshauer, G.E. and McCourt, M. *Kernel-based approximation methods using MATLAB*, Interdisciplinary Mathematical Sciences, Vol. 19, World Scientific Publishing Co., Singapore, 2015.

[9] Fornberg, B. and Flyer, N. *A primer on radial basis functions with applications to the geosciences*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2015.

[10] Fornberg, B. and Flyer, N. *Solving PDEs with radial basis functions,* Acta Numer. 24 (2015), 215–258.

[11] Fornberg, B., Larsson, E. and Flyer, N. Stable computations with Gaussian radial basis functions, SIAM J. Sci. Comput. 33(2) (2011), 869–892.

[12] Fornberg, B. and Wright, G. *Stable computation of multiquadric interpolants for all values of the shape parameter,* Comput. Math. Appl. 48(5–6) (2004), 853–867.

[13] Garmanjani, G., Banei, S., Shanazari, K. and Azari, Y. *An RBF-PUM finite difference scheme for forward–backward heat equation,* Comput. Appl. Math. 42 (2023), 231.

[14] Hussain, M., Ghafoor, A., Hussain, A., Haq, S., Ali, I. and Ul Arifeen, S. *A hybrid kernel-based meshless method for numerical approximation of multidimensional Fisher's equation,* Math. Comput. Simul. 223 (2024), 130–157.

[15] Larsson, E., Shcherbakov, V. and Heryudono, A. *A least squares radial basis function partition of unity method for solving PDEs,* SIAM J. Sci. Comput. 39(6) (2017), A2539–A2564.

[16] Melenk, J.M. and Babuška, I. *The partition of unity finite element method: Basic theory and applications,* Comput. Methods Appl. Mech. Eng. 139(1) (1996), 289–314.

[17] Mirzaei, D. *The direct radial basis function partition of unity (D-RBF-PU) method for solving PDEs,* SIAM J. Sci. Comput. 43(1) (2021), A1223–A1249.

[18] Mishra, P.K., Fasshauer, G.E., Sen, M.K. and Ling, L. *A stabilized radial basis-finite difference (RBF-FD) method with hybrid kernels,* Comput. Math. Appl. 77(9) (2019), 2354–2368.

[19] Mohammadi, M., Sommariva, A. and Vianello, M. *Unisolvence of Kansa collocation for elliptic equations by polyharmonic splines with random fictitious centers,* Appl. Math. Lett. 168 (2025), 109571.

[20] Wendland, H. *Fast evaluation of radial basis functions: Methods based on partition of unity*, in Approximation Theory X: Wavelets, Splines, and Applications, Vanderbilt University Press, Nashville, 2002, 473–483.

[21] Wendland, H. *Scattered Data Approximation*, Cambridge University Press, 2005.

Iranian Journal of Numerical Analysis and Optimization
Vol. 17, No. 1, 2027, pp 72–108
https://doi.org/10.22067/ijnao.2026.96859.1794
https://ijnao.um.ac.ir/

CC BY NC

Research Article

Behavioral modeling and control strategies for Rumor-induced customer dynamics on digital platforms

S. Belhdid*, H. Hassouni, O. Aarabate and O. Balatif

Abstract

This paper proposes a compartmental mathematical model for analyzing rumor-induced behavioral dynamics among customers on digital platforms. The customer population is divided into three the model captures how misinformation influences transitions between these states. The positivity and boundedness of the solutions are established, and the basic reproduction

*Corresponding author

Received 5 December 2025; revised 13 May 2026; accepted 5 August 2026

Salaheddine Belhdid
Laboratory of Fundamental Mathematics and their Applications, Department of Mathematics, Faculty of Sciences El Jadida, Chouaib Doukkali University, El Jadida, Morocco. e-mail: salaheddine.belhdid@gmail.com

Habib Hassouni
Laboratory of Fundamental Mathematics and their Applications, Department of Mathematics, Faculty of Sciences El Jadida, Chouaib Doukkali University, El Jadida, Morocco. e-mail: habibhassouni85@gmail.com

Oumaima Aarabate
Laboratory of Fundamental Mathematics and their Applications, Department of Mathematics, Faculty of Sciences El Jadida, Chouaib Doukkali University, El Jadida, Morocco. e-mail: oumaimaaarabate@gmail.com

Omar Balatif
Laboratory of Fundamental Mathematics and their Applications, Department of Mathematics, Faculty of Sciences El Jadida, Chouaib Doukkali University, El Jadida, Morocco. e-mail: balatif.maths@gmail.com

How to cite this article
Belhdid, S., Hassouni, H., Aarabate, O. and Balatif, O., Behavioral modeling and control strategies for Rumor-induced customer dynamics on digital platforms. *Iran. J. Numer. Anal. Optim.*, 2027; 17(1): 72-108. https://doi.org/10.22067/ijnao.2026.96859.1794

number $\mathcal{R}_0$ is derived via the next-generation matrix method and is used to characterize the stability of the rumor-free and endemic equilibria. Both local and global stability analyses are carried out using linearization and Lyapunov methods, respectively. An optimal control problem is then formulated with two time-dependent intervention strategies: Incentive-based and support-based controls. The existence of optimal controls is verified, and the necessary optimality conditions are derived via Pontryagin's maximum principle. Numerical simulations based on a forward-backward sweep scheme illustrate the effectiveness of the proposed strategies in reducing hesitation and enhancing long-term customer loyalty, with combined interventions yielding the best results.

AMS subject classifications (2020): Primary 49J20; Secondary 34D20.

Keywords: Compartmental models; Optimal control; Rumors; Customer behavior; Numerical simulations; Digital platforms.

1 Introduction

Rumor dissemination on digital platforms profoundly influences customer behavior, shaping perceptions, trust, and engagement within online ecosystems. In today's hyperconnected environment, false or misleading information can spread rapidly, much like infectious diseases, thereby affecting how potential users transition toward or away from loyalty. Understanding and mitigating these rumor-induced behavioral dynamics have become essential for maintaining the sustainability and credibility of digital platforms.

This study investigates the interplay between rumor propagation and customer behavioral dynamics through a compartmental mathematical model that distinguishes among potential, hesitant, and loyal customers. Drawing from epidemiological modeling frameworks, we analyze how information diffusion, specifically the spread of tumor, affects customer conversion, hesitation, and retention processes [18]. By formulating the problem in this manner, the model captures the contagion-like properties of rumor dissemination and enables the development of optimal control strategies to mitigate its adverse effects while enhancing user engagement [20, 22].

The theoretical foundation of this research is based on classical and contemporary rumor spreading models, which treat the dissemination of information as analogous to the spread of infectious diseases. These models have evolved to incorporate key behavioral and cognitive factors such as education, debunking efforts, and intervention strategies, highlighting their role in reducing misinformation persistence [18, 21]. Several studies have also accounted for individual heterogeneity, distinguishing among those who propagate rumors, those who identify or refute

them, and those who remain indifferent, thereby offering a more realistic portrayal of social communication processes [18, 14].

Recent advances in compartmental modeling and optimal control have highlighted their utility in analyzing complex social and behavioral dynamics. In particular, the application of SIR models to information and rumor diffusion has grown substantially [18, 6, 25, 27], with several studies demonstrating the importance of control-theoretic approaches in managing misinformation [9, 4, 19]. These works collectively establish the mathematical foundation on which the present study builds, while our contribution lies in extending this framework to a customer behavioral context with an explicit hesitancy compartment. For example, El Bhih et al. [9] developed a framework to study the spread of rumors and counter-rumors on social networks, formulating multiple equilibria and designing optimal interventions using Pontryagin's maximum principle to influence collective behavior. Similarly, studies on scam rumor propagation have employed deterministic models combined with cost-effectiveness analysis to evaluate strategies for mitigating false information on social platforms, emphasizing the efficiency of targeted interventions [4]. In the context of epidemiology, Laaroussi, El Bhih, and Rachik [15] demonstrated how optimal vaccination and treatment policies can be formulated under constrained resources for a multi-strain SEIR model of COVID-19, providing insight into balancing limited resources while achieving maximal impact. Furthermore, Yaagoub, El Bhih, and Allali [26] introduced a fractional-order infection model for computer virus propagation, illustrating how memory effects and fractional dynamics can improve the realism and predictive power of compartmental models. Collectively, these studies underscore the versatility of compartmental and optimal control approaches for understanding the evolution of complex systems and designing effective intervention strategies. Inspired by these works, our study applies a similar framework to investigate merchants' adoption of e-commerce, where the dynamics of acceptance and refusal can be influenced through targeted control measures. However, most existing works rely on simplified tow-state models, typically classifying individuals as either active or inactive spreaders of information [12]. While analytically tractable, such models overlook the behavioral nuances and decision-making diversity inherent in customer interactions on digital platforms. In contrast, recent studies emphasize that users exhibit varying degrees of commitment and susceptibility ranging from potential adopters who are exploring the platform to hesitant users influenced by rumors, to loyal customers who reinforce platform stability [11, 2].

The present work addresses this gap by proposing a multi-compartmental behavioral model that explicitly differentiates these customer categories and incorporates the effect of rumor propagation on their transitions. This approach aligns with recent advances in information diffusion and population dynamics modeling, which account for delays, cross-transmission, and endogenous attention mechanisms [3, 7]. Furthermore, the model recognizes that algorithmic recom-

mendation systems can inadvertently amplify misinformation by promoting highly engaging but unreliable content, thereby reinforcing hesitancy among potential users [13].

In competitive digital environments, media competition and attention-driven algorithms may unintentionally fuel the spread of misinformation, undermining both user trust and long-term platform growth [1]. To confront these challenges, our study integrates behavioral modeling, rumor propagation dynamics, and optimal control theory into a single mathematical framework. By doing so, it seeks to analyze the stability properties of the system and design effective control strategies such as rumor refutation campaigns or customer education initiatives that can enhance platform resilience.

Ultimately, this research contributes to the growing body of work that applies mathematical modeling and control theory to social and behavioral systems. By coupling rumor dynamics with customer segmentation, the proposed model provides a deeper understanding of how misinformation influences user transitions and engagement. The resulting insights offer both theoretical and practical guidance for platform managers and policymakers aiming to balance information flow, customer loyalty, and sustainable growth in the digital era [8].

The proposed model offers three key advantages over existing approaches. First, unlike classical two-state rumor models that classify individuals simply as spreaders or ignorant, our model captures the intermediate *hesitant* state, which is empirically more realistic for customer behavior on digital platforms where indecision is prevalent. Second, the use of frequency-dependent interaction rather than mass-action kinetics ensures that the model remains well posed as the total population varies over time. Third, the integration of two distinct and complementary optimal control strategies: incentive-based and support-based allows for a richer intervention design than a single-control formulation, while Pontryagin's maximum principle provides a rigorous and tractable characterization of optimal policies.

The remainder of the paper is structured as follows. Section 2 introduces the proposed multi-compartmental model, detailing the dynamics of potential, hesitant, and loyal customers and the mechanisms of rumor propagation. Section 3 is dedicated to the mathematical analysis of the model, including the study of equilibrium points, stability properties, and the basic reproduction number. Section 4 formulates the optimal control problem, defining admissible strategies, objective functionals, and the associated existence results. Additionally, it drives the necessary conditions for optimality using Pontryagin's maximum principle and presents the characterization of the optimal controls. Section 5 provides numerical simulations to illustrate the effectiveness of the control strategies under various scenarios and sensitivity analyses of key parameters. Finally, Section 6 concludes the paper with a discussion of the main findings, practical implications, and potential directions for future research.

2 Mathematical model overview:

The dynamical system under consideration describes the interaction and engagement dynamics of users on a digital platform. Three distinct compartments encapsulate the diverse user categories: Potential customers ($P(t)$), loyal customers ($L(t)$), and hesitant customers ($H(t)$). Each compartment represents a specific phase in the user journey, from potential consideration to active engagement or hesitancy. The model accounts for the interplay between these compartments, considering factors such as attraction, conversion, retention, and natural attrition. This detailed analysis aims to shed light on the complex dynamics influencing user engagement, offering valuable insights for optimizing and enhancing the digital platform's user experience. The model variables are summarized in Table 1.

Table 1: Description of the model variables

$P(t)$	Potential customers or users considering the digital platform.
$L(t)$	Loyal customers who actively engage with the digital platform.
$H(t)$	Hesitant customers who are interested but not fully engaged.

Potential customers ($P(t)$):

This compartment represents individuals who are in the consideration phase, pondering the use of the digital platform. Attraction occurs at a rate Ω_1, reflecting the platform's appeal. Interactions play a crucial role, with potential customers converting to loyal customers ($\beta_1 P(t)L(t)$) or hesitant customers ($\beta_2 P(t)H(t)$), influenced by the existing user base. Natural attrition (μ) accounts for potential customers naturally disengaging. The overall dynamic process involves a delicate balance between attraction, conversion to loyalty, conversion to hesitancy, and the inherent attrition rate.

Loyal customers ($L(t)$):

This compartment represents actively engaged users who have transitioned from potential customers. The dynamic process is shaped by the conversion of potential customers to loyal customers ($\beta_1 P(t)L(t)$), their own retention (α_1), and the conversion of loyal customers to hesitant ones (π). Retention (α_1) captures the platform's ability to retain potential customers as loyal users. The conversion dynamics involve both attracting new users and converting loyal users to hesitant ones. Natural attrition (μ) also affects this group, reflecting the natural disengagement of loyal users from the platform.

Hesitant customers ($H(t)$):

This compartment represents individuals intrigued by the platform but not fully committed. The dynamic process involves potential customers converting to hesitant customers ($\beta_2 P(t)H(t)$), their own retention (α_2), and the conversion of hesitant customers to loyal ones (τ). Retention (α_2) signifies the platform's ability to retain potential customers as hesitant users. The conversion dynamics reflect the interplay between attracting new hesitant users and converting hesitant users to loyal ones. Natural attrition (μ) affects hesitant customers similarly to loyal customers. The overall dynamic process encompasses attraction, conversion to loyalty, conversion to hesitancy, and the inherent attrition rate, contributing to a comprehensive analysis of user engagement on the digital platform. The corresponding compartmental flow diagram is depicted in Figure 1, and the resulting dynamics are governed by the following system of ordinary differential equations:

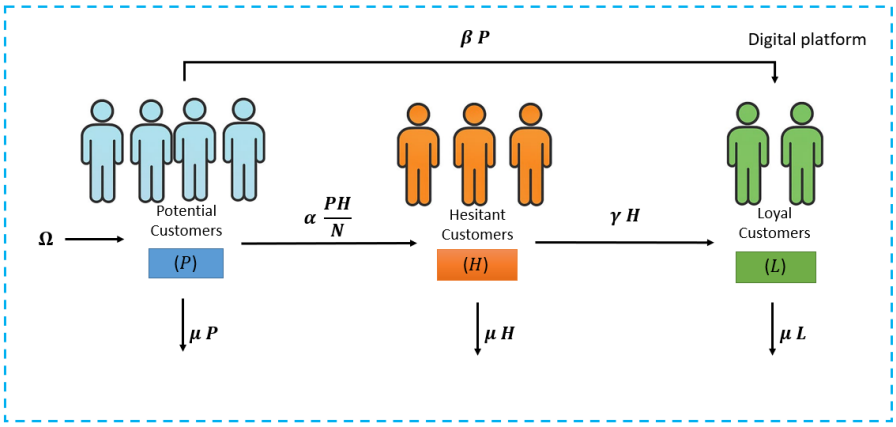


Figure 1: The compartmental flow diagram for the customers spreading process.

$$\begin{aligned}
\frac{dP(t)}{dt} &= \Omega - \alpha\frac{P(t)H(t)}{N} - (\beta+\mu)P(t),\\
\frac{dH(t)}{dt} &= \alpha\frac{P(t)H(t)}{N} - (\gamma+\mu)H(t),\\
\frac{dL(t)}{dt} &= \beta P(t) + \gamma H(t) - \mu L(t),
\end{aligned} \tag{1}$$

with the initial conditions

$$P(0) \geq 0, \quad L(0) \geq 0, \quad H(0) \geq 0.$$

Parameters	Meaning
Ω	Attraction or onboarding rate of potential customers to the platform.
β	Conversion rate from potential to loyal customers.
α	Conversion rate from potential to hesitant customers.
μ	Natural loss rate of customers (attrition rate).
γ	Rate at which hesitant customers become loyal customers.

3 Model analysis

3.1 Positivity and boundedness

Considering the model system (1), we provide the following:

$$\begin{aligned}
\left.\frac{dP}{dt}\right|_{\{P=0,\ H>0,\ L>0\}} &= \Omega > 0,\\
\left.\frac{dH}{dt}\right|_{\{P>0,\ H=0,\ L>0\}} &= 0,\\
\left.\frac{dL}{dt}\right|_{\{P>0,\ H>0,\ L=0\}} &= \beta P + \gamma H > 0.
\end{aligned}$$

Observing that all rates are nonnegative at the boundary of the nonnegative cone in $\mathbb{R}^3$, and considering the inward directions of the vector fields on these boundaries, initiating the system from any interior point of this cone guarantees that the solutions will persist within the cone indefinitely. This guarantees the positivity of all populations in the system over time.

Theorem 1. The set $\Theta = \{(P, H, L) \in \mathbb{R}^3_+ : 0 \leq P + L + H \leq \frac{\Omega}{\mu}\}$ with the initial conditions $P(0) \geq 0,\ L(0) \geq 0,\ H(0) \geq 0$, is a positively invariant and global attractive set under the system (1).

Proof. Note that the total population N is given by $N = P + L + H$. Differentiating, we obtain that

$$\frac{dN}{dt} = \Omega - \mu N,$$

thereafter, by solving the rate equation we get the solution $N(t)$ as follows:

$$N(t) = \frac{\Omega}{\mu} - \left(\frac{\Omega}{\mu} + N(0)\right) \exp^{-\mu t}. \tag{2}$$

If $N(t) > \frac{\Omega}{\mu}$, then from (2), we observe that $\lim_{t\to\infty} N(t) = \frac{\Omega}{\mu}$. If $N(0) \leq \frac{\Omega}{\mu}$, then $N(t) \leq \frac{\Omega}{\mu}$. This shows that all solutions with initial conditions in Θ stay in Θ for all future times, and consequently the set Θ is a positively invariant for the system (1). Therefore, we get to the conclusion that the set Θ is a positively invariant set and globally attractive for the model system (1).

□

Remark 1. Adding the three equations of system (1), the total population $N(t) = P(t)+H(t)+L(t)$ satisfies

$$\frac{dN}{dt} = \Omega - \mu N(t),$$

whose explicit solution is

$$N(t) = \frac{\Omega}{\mu} - \left(\frac{\Omega}{\mu} - N(0)\right) e^{-\mu t}.$$

Therefore, $N(t)$ is not constant in general; it varies over time unless $N(0) = \Omega/\mu$ exactly. However, $N(t)$ is bounded and converges asymptotically to Ω/μ as $t \to \infty$, regardless of $N(0)$. This justifies the use of the time-varying $N(t)$ in the frequency-dependent interaction term $\alpha P(t)H(t)/N$, which normalizes the rumor contact rate by the current total population, consistent with standard compartmental modeling conventions. In the numerical simulations, $N(t)$ is updated at each time step of the discretization scheme.

3.2 Existence of free-equilibrium point and Basic reproduction number

The free-equilibrium point is computed by equating the right-hand side of the system (1) to zero, that is,

$$\frac{dP(t)}{dt} = 0, \quad \frac{dH(t)}{dt} = 0, \quad \frac{dL(t)}{dt} = 0. \tag{3}$$

Therefore, the free-equilibrium steady state of the model (1) is the state in which there are no loyal customers and no hesitant customers into the total population, that is,

$$H(t) = 0, \quad L(t) = 0.$$

Thus, after equating the right-hand side of the system (1) to zero and substituting these values into the system, we get

$$\Omega - (\beta + \mu)P = 0 \quad \Longrightarrow \; P^* = \frac{\Omega}{\beta + \mu},$$

and

$$\beta P - \mu L = 0 \quad \Longrightarrow \; L^* = \frac{\beta \Omega}{\mu(\beta + \mu)}.$$

Consequently, the free-equilibrium $\mathcal{E}_f$ is given by

$$\mathcal{E}_f = \left\{ (P^*, 0, L^*) : \left(\frac{\Omega}{\beta + \mu}, 0, \frac{\beta \Omega}{\mu(\beta + \mu)} \right) \right\}, \tag{4}$$

which always exists.

We now derive the basic reproduction number $\mathcal{R}_0$ of the model (1), given by the following theorem.

Theorem 2. The basic reproduction number of the model system (1) is stated as follows:

$$\mathcal{R}_0 = \alpha \frac{\mu}{(\beta + \mu)(\gamma + \mu)}, \tag{5}$$

Proof. Use the next generation matrix [24]. Let $X = (P, H, L)$; then system can be written

$$\frac{dX}{dt} = f(X) - v(X),$$

where

$$f(X) = \begin{pmatrix} \frac{\alpha P H}{N} \\ 0 \\ 0 \end{pmatrix},$$

and

$$v(X) = \begin{pmatrix} (\gamma + \mu) H \\ -\beta P - \gamma H + \mu L \\ -\Omega + \frac{\alpha P H}{N} + (\beta + \mu) P \end{pmatrix}.$$

Then, the two Jacobian matrices at the free-equilibrium point for $f(X)$ and $v(X)$ are written as follows:

$$J\left(f\left(\mathcal{E}_f\right)\right) = \begin{pmatrix} F_{2\times 2} & 0 \\ & 0 \\ 0 \quad 0 & 0 \end{pmatrix},$$

and

$$J\left(v\left(\mathcal{E}_f\right)\right) = \begin{pmatrix} V_{2\times 2} & 0 \\ & -\beta \\ \dfrac{\mu\alpha}{(\beta+\mu)} \; 0 & 0 \end{pmatrix},$$

with

$$F_{2\times 2} = \begin{pmatrix} \alpha\frac{\mu}{\beta+\mu} & 0 \\ 0 & 0 \end{pmatrix},$$

and

$$V_{2\times 2} = \begin{pmatrix} (\gamma+\mu) & 0 \\ -\gamma & \mu \end{pmatrix}.$$

Thereafter, by a simple calculation we get the following:

$$\mathcal{R}_0 = \rho(F_{2\times 2}V_{2\times 2}^{-1}) = \alpha\frac{\mu}{(\beta+\mu)(\gamma+\mu)}.$$

□

Remark 2. In the context of the compartmental model describing customer behavior on digital platforms, the basic reproduction number $\mathcal{R}_0$ represents the expected number of new hesitant customers generated by one hesitant customer in the absence of other influences. Specifically, $\mathcal{R}_0$ is a threshold parameter that helps determine the potential for customer base growth. If $\mathcal{R}_0 > 1$, each potential customer, on average, leads to more than one hesitant customer, indicating that the platform's customer base is likely to grow. Conversely, if $\mathcal{R}_0 < 1$, the platform struggles to convert potential customers into hesitant ones at a sufficient rate, suggesting that the customer base may decline over time. Understanding $\mathcal{R}_0$ helps the platform strategize its marketing and engagement efforts to ensure sustainable growth and transition towards a stable, loyal customer base.

3.3 Sensitivity analysis in Rumor-induced customer dynamics

Sensitivity analysis serves as a pivotal tool to identify and quantify the impact of key model parameters on the prevalence of hesitant customers and the overall influence of rumors within digital platforms. In the context of our study, the basic reproduction number R_0 plays a crucial role, characterizing the potential for rumor-induced transitions among customer segments. Understanding how variations in model parameters affect R_0 provides insights into the dynamics of potential, hesitant, and loyal customers, thereby informing strategies to mitigate the adverse effects of misinformation [6, 19].

In this section, we examine the effect of each parameter in system (1) on the value of R_0. To determine which parameters most significantly drive the spread of rumor-induced hesitation and affect customer conversion to loyalty, a sensitivity analysis is conducted following the approaches in [25, 27]. Specifically, we compute the partial derivatives of R_0 with respect to the system parameters in (1) to obtain the corresponding sensitivity indices (S.Is.), providing a quantitative measure of each parameter's influence on the dynamics of potential, hesitant, and loyal customers.

This analysis not only enhances our understanding of the model's behavior but also aids in identifying critical parameters that can be targeted to control the spread of rumors and promote customer loyalty. By focusing on these influential parameters, digital platforms can develop more effective strategies to manage misinformation and its impact on user engagement [19, 6].

Definition 1. The S.I. $Z_q^{R_0}$ of the basic reproduction number R_0, varying with parameter q, is defined as

$$Z_q^{R_0} = \frac{\partial R_0}{\partial q} \cdot \frac{q}{R_0},$$

where $\frac{\partial R_0}{\partial q}$ denotes the partial derivative of R_0 with respect to q.

Specifically, the computed S.Is. of the basic reproduction number R_0 with respect to the parameters of the controlled PHL system are as follows:

$$\begin{cases} \Upsilon_{\mu}^{R_0} = \dfrac{\partial R_0}{\partial \mu} \cdot \dfrac{\mu}{R_0} = \dfrac{\gamma\beta - \mu^2}{(\gamma+\mu)(\beta+\mu)}, \\ \Upsilon_{\alpha}^{R_0} = \dfrac{\partial R_0}{\partial \alpha} \cdot \dfrac{\alpha}{R_0} = 1, \\ \Upsilon_{\beta}^{R_0} = \dfrac{\partial R_0}{\partial \beta} \cdot \dfrac{\beta}{R_0} = -\dfrac{\beta}{\beta+\mu}, \\ \Upsilon_{\gamma}^{R_0} = \dfrac{\partial R_0}{\partial \gamma} \cdot \dfrac{\gamma}{R_0} = -\dfrac{\gamma}{\gamma+\mu}. \end{cases}$$

These indices quantify the relative impact of each parameter on the reproduction number, providing insight into which factors most strongly influence the spread of hesitant (H) and loyal (L) customer behaviors. The corresponding numerical values, computed at the nominal parameter set, are reported in Table 2. These results are visualized in Figure 2, which displays the S.Is. of R_0 with respect to each parameter.

Table 2: Sensitivity of the basic reproduction number R_0 to parameter perturbations (nominal values: $\alpha = 5.0 \times 10^{-4}$, $\beta = 1.0 \times 10^{-4}$, $\gamma = 5.0 \times 10^{-6}$, $\mu = 3.0 \times 10^{-4}$).

Parameter	Nominal value	S.I. $Z_q^{R_0}$	R_0 (+10%)	ΔR_0 (+10%) (percent)
α	5.00×10^{-4}	1.0000	1.3525	+10.00%
β	1.00×10^{-4}	−0.2500	1.1995	−2.44%
γ	5.00×10^{-6}	−0.01639	1.2275	−0.16%
μ	3.00×10^{-4}	−0.73361	1.1454	−6.84%

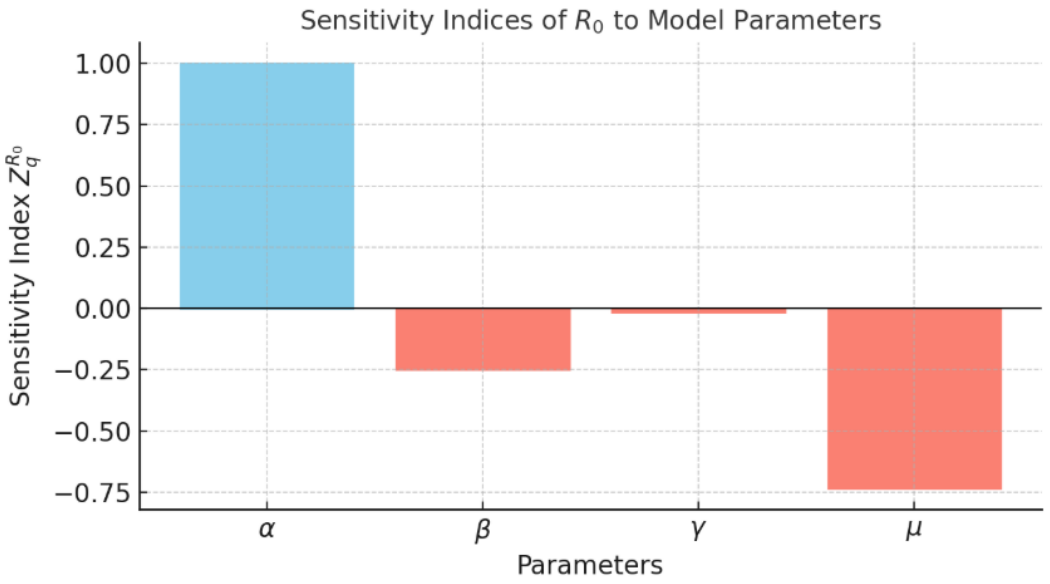


Figure 2: Sensitivity analysis of the parameters.

3.4 Existence of an endemic steady state

The model system (1) admits *two* distinct equilibrium points: The rumor-free equilibrium $\mathcal{E}_f$ (established in Section 3.2), which exists for all parameter values, and the endemic equilibrium E_e^* described below, which exists and is biologically meaningful only when $\mathcal{R}_0 > 1$.

The endemic equilibrium point in this customer behavior model toward digital platforms represents a stable state where the proportions of potential, hesitant, and loyal customers remain constant over time. At this point, the rates of transitions between the compartments (potential to hesitant, hesitant to loyal, etc.) balance each other out, resulting in no net changes in the number of customers in each compartment. This equilibrium reflects a situation where the digital platform has reached a steady state in customer dynamics, with consistent engagement and retention patterns. Understanding this equilibrium can help the platform predict long-term customer behavior and design strategies to maintain or shift this balance in the desired directions. Therefore, the endemic-steady state for our model is given by the following:

$$E_e^* = (P^*, H^*, L^*)\,,$$

where

$$P^* = \Omega \frac{(\gamma+\mu)}{\mu\alpha},$$

$$H^* = \Omega \frac{(\beta+\mu)(\mathcal{R}_0-1)}{\mu\alpha},$$

and

$$L^* = \frac{\beta\Omega^2(\gamma+\mu)^2 + \gamma\Omega(\beta+\mu)(\gamma+\mu)}{\mu\alpha(\gamma+\mu)}(\mathcal{R}_0-1).$$

3.5 Local stability analysis of the steady states

Theorem 3. Considering the system model (1) at the free-equilibrium $\mathcal{E}_f$, it is locally asymptotically stable if $\mathcal{R}_0 < 1$, and unstable when $\mathcal{R}_0 > 1$.

Proof. We compute the Jacobian of system (1) evaluated at $\mathcal{E}_f$:

$$J(\mathcal{E}_f) = \begin{bmatrix} -\beta-\mu & -\alpha\frac{\mu}{\beta+\mu} & 0 \\ 0 & (R_0-1)(\gamma+\mu) & 0 \\ \beta & \gamma & -\mu \end{bmatrix}.$$

Therefore, the eigenvalues of the characteristic equation of $J(\mathcal{E}_f)$ are
$v_1 = (R_0-1)(\gamma+\mu) < 0,$
$v_2 = -\mu < 0,$
$v_3 = -\beta-\mu < 0.$

Therefore, all the eigenvalues of the characteristic equation are negative if $R_0 < 1$. Hence, the equilibrium point $\mathcal{E}_f$ is locally asymptotically stable if $R_0 < 1$ and unstable if $R_0 > 1$.

□

Theorem 4. If $R_0 > 1$, the endemic equilibrium E_e^* is locally asymptotically stable.

Proof. Let us consider the endemic equilibrium point E_e^*, given by

$$X_* = \left(\frac{\Omega}{\alpha\mu}(\gamma+\mu), \left(\frac{\Omega}{\alpha\mu}(R_0-1)(\beta+\mu)\right), \frac{1}{\alpha\mu^2}(\Omega\beta\mu + \Omega\beta\gamma R_0 + \Omega\gamma\mu(R_0-1))\right).$$

Then the Jacobian matrix at E_e^* is given as follows:

$$J(E_e^*) = \begin{bmatrix} -\alpha\frac{H^*}{N} - (\beta+\mu) & -\alpha\frac{P^*}{N} & 0 \\ \alpha\frac{H^*}{N} & \alpha\frac{P^*}{N} - (\gamma+\mu) & 0 \\ \beta & \gamma & -\mu \end{bmatrix},$$

and its characteristic equation is

$$X^3 + a_2X^2 + a_1X + a_0 = 0,$$

where

$$a_2 = \frac{\mu\alpha}{(\beta+\mu)} + \mu,$$
$$a_1 = \frac{\mu^2\alpha}{(\beta+\mu)} + (\beta+\mu)(\gamma+\mu)\,(R_0-1)\,,$$
$$a_0 = \frac{\mu^3\alpha}{(\beta+\mu)N} + \mu(\beta+\mu)(\gamma+\mu)\,(R_0-1)\,.$$

By the Routh–Hurwitz criterion, system (1) is locally asymptotically stable if $a_2 > 0$, $a_0 > 0$, and $a_2a_1 > a_0$. Thus, E_e^* is locally asymptotically stable if $R_0 > 1$. □

3.6 Global stability analysis

Now, we are concerned with the global asymptotic stability of free equilibrium $\mathcal{E}_f$ and endemic equilibrium E_e^* of model (1), respectively.

Theorem 5. The free equilibrium $\mathcal{E}_f$ is globally asymptotically stable on Θ, if $R_0 \leq 1$.

Proof. To prove the global stability of the free equilibrium $\mathcal{E}_f$, we consider the Lyapunov function constructed as follows: $V : \Omega \to \mathbb{R}$ given by

$$V(P,H) = \frac{1}{2}\left((P-P_0)+H\right)^2 + \frac{(\beta+\gamma+2\mu)}{\alpha}NH.$$

Then, the time derivative of V is

$$\begin{aligned}
V'(P,H) &= \big((P-P_0)+H\big)\big(P'+H'\big) + \frac{\beta+\gamma+2\mu}{\alpha}NH', \\
&= \left(\left(P - \frac{\Omega}{\beta+\mu} + H\right)\left(\Omega - \frac{\alpha PH}{N} - (\beta+\mu)P + \frac{\alpha PH}{N} - (\gamma+\mu)H\right)\right. \\
&\quad \left. + \frac{\beta+\gamma+2\mu}{\alpha}N\left(\frac{\alpha PH}{N} - (\gamma+\mu)H\right)\right), \\
&= -(\beta+\mu)\left(P - \frac{\Omega}{\beta+\mu}\right)^2 - (\gamma+\mu)H^2 \\
&\quad - (\beta+\gamma+2\mu)\frac{\gamma+\mu}{\alpha}N(1-R_0)H.
\end{aligned}$$

Thus, $V^{'}(P,H) \leq 0$ for $R_0 \leq 1$.
In addition, If $R_0 \leq 1$, then $V'(P,H) = 0 \iff P = P_0$ and $H = 0$.

Hence, by LaSalle's invariance principle [16], the free equilibrium point $\mathcal{E}_f$ is globally asymptotically stable on Θ. □

Theorem 6. The endemic equilibrium E_e^* is globally asymptotically stable on Θ, if $R_0 > 1$.

Proof. To prove the global stability of the endemic equilibrium E_e^*, we consider the Lyapunov function constructed as follows: $V : \Omega \to \mathbb{R}$ given by

$$V(P,H) = C_1\left(P - P^* \ln\frac{P}{P^*}\right) + C_2\left(H - H^* \ln\frac{H}{H^*}\right),$$

where C_1 and C_2 are positive constants to be chosen later. Then, the time derivative of the Lyapunov function is given by

$$V^{'}(P,H) = \frac{\alpha}{N}(C_2 - C_1)(H - H^*)(P - P^*) - \Omega C_1\frac{(P-P^*)^2}{PP^*}.$$

For $C_1 = C_2 = 1$, we get

$$V^{'}(P,H) = -\Omega\frac{(P-P^*)^2}{PP^*} \leq 0,$$

we also get

It remains to identify the largest invariant set contained in $\mathcal{M} = \{(P,H) \in \Theta : V'(P,H) = 0\} = \{(P,H) \in \Theta : P = P^*\}$. On this set, $\dot{P} = 0$, so the first equation of system (1) gives $\Omega - \alpha P^* H/N - (\beta+\mu)P^* = 0$. Since at the endemic equilibrium $\Omega = \alpha P^* H^*/N + (\beta+\mu)P^*$, substituting yields $\alpha P^*(H - H^*)/N = 0$, and hence $H = H^*$. Therefore, the largest invariant subset of $\mathcal{M}$ is the singleton $\{E_e^*\}$.

$V^{'}(P,H) = 0 \iff P = P^*$.

Hence, by LaSalle's invariance principle [16], the endemic equilibrium point E_e^* is globally asymptotically stable on Θ. □

4 Optimal control problem

The current section is devoted to the model's optimal control analysis. By implementing two time-dependent controls $v(t)$ and $w(t)$ to our proposed system (1), we obtain the following controlled dynamical system governed by the upcoming differential equations:

$$\frac{dP(t)}{dt} = \Omega - \alpha\frac{P(t)H(t)}{N} - (\beta+\mu)P(t) - v(t)P(t),$$

$$\begin{aligned}\frac{dH(t)}{dt} &= \alpha\frac{P(t)H(t)}{N} - (\gamma+\mu)H(t) - w(t)H(t),\\ \frac{dL(t)}{dt} &= \beta P(t) + \gamma H(t) - \mu L(t) + v(t)P(t) + w(t)H(t),\end{aligned} \tag{6}$$

with the initial conditions

$$P(0) \geq 0, \quad L(0) \geq 0, \quad H(0) \geq 0.$$

Taking into consideration the economic and ethical restrictions of the digital platforms, the intervention policies must be restricted. Consequently, the two controls have to be taken as bounded in $[0,1]$. The control strategies $v(t)$ and $w(t)$ are relevant precisely when $\mathcal{R}_0 > 1$, that is, when the rumor-free equilibrium $\mathcal{E}_f$ is unstable and the system tends toward the endemic equilibrium E_e^*. When $\mathcal{R}_0 \leq 1$, the uncontrolled system already converges to $\mathcal{E}_f$ (see Theorem 5), and no intervention is required. The optimal controls are therefore designed to drive the system back toward a low-hesitancy state when the reproduction number exceeds the critical threshold.

1. **Incentive-based control strategy** $v(t)$ is a strategy used in various fields, such as economics, environmental management, and public health, to influence the behavior of individuals or groups by offering rewards or penalties. The underlying idea is to align individual motivations with broader societal goals by creating incentives that encourage desirable actions while discouraging harmful ones. For example, in environmental management, incentives like tax credits or subsidies may be offered to companies that reduce pollution, while fines or higher taxes might be imposed on those that exceed pollution limits. In public health, similar strategies can be applied, such as offering financial incentives to people who engage in healthy behaviors or penalizing risky behaviors that contribute to public health issues. Incentive-based control leverages human motivation to drive change without relying solely on direct regulations or restrictions, allowing for flexibility and innovation in achieving desired outcomes.

2. **Support-based control strategy** $w(t)$ focuses on providing resources, tools, or assistance to individuals or organizations to help them achieve desired behaviors or outcomes. Rather than using incentives or penalties, these strategies aim to empower and enable change through guidance, education, or infrastructural support. In environmental management, for instance, support-based strategies might involve offering technical assistance or providing access to renewable energy technologies to help businesses reduce emissions. In public health, these strategies could include offering free healthcare services, counseling, or educational programs to promote healthier lifestyles. By addressing barriers to positive behavior, such as lack of knowledge, resources, or access support-based control strategies

foster an environment where individuals and groups can make sustainable changes. This approach tends to be collaborative and often works best when paired with other control mechanisms, like regulations or incentive-based programs, to encourage long-term compliance and growth.

In summary, this section has formulated the optimal control problem for the proposed PHL system by introducing two time-dependent interventions: The incentive-based control $v(t)$ and the support-based control $w(t)$. Both controls are bounded to respect practical and ethical constraints, reflecting realistic implementation scenarios. The incentive-based strategy targets behavior modification through rewards or penalties, while the support-based strategy provides resources and assistance to facilitate positive outcomes. Together, these controls establish a flexible and practical framework for designing intervention policies that aim to optimize system performance while adhering to societal and economic considerations. This formulation lays the foundation for the subsequent analysis, including the existence of optimal controls, derivation of necessary conditions, and numerical simulations to evaluate their effectiveness.

4.1 Objective functional

We aim to minimize the following objective functional:

$$J(v,w) = H(t_f) - L(t_f) + \int_{t_0}^{t_f} \Big[H(t) - L(t) + \frac{M_1}{2}v^2(t) + \frac{M_2}{2}w^2(t)\Big]dt, \tag{7}$$

where t_f is the final time and $M_1, M_2 > 0$ are cost coefficients that weigh the relative importance of the controls v and w.

We search for optimal controls (v^*, w^*) such that

$$J(v^*, w^*) = \min_{(v,w)\in U_{ad}^2} J(v,w),$$

with the admissible control set

$$U_{ad} = \{u(t) : 0 \le u(t) \le 1,\ t \in [t_0, t_f]\}.$$

From a broader perspective, the objective functional is designed to manage the trade-off between converting hesitant customers, represented by $H(t)$, into more loyal behavior, and preserving the existing loyal customer base, represented by $L(t)$. At the same time, it incorporates the

cost of implementing the incentive-based and support-based control strategies $v(t)$ and $w(t)$ through the weighting coefficients M_1 and M_2, ensuring that interventions remain both effective and feasible within practical economic constraints. By restricting the admissible controls to the interval $[0, 1]$, we reflect realistic limitations on the intensity of policy implementation. This formulation provides a rigorous approach for deriving optimal strategies that encourage hesitant customers to engage positively with the system while safeguarding the loyalty of established customers. Moreover, it sets the foundation for numerical simulations to evaluate the impact of different combinations of incentives and support measures on customer dynamics and overall system performance.

4.2 Existence of optimal controls

Fleming and Rishel's result (see [10, Corollary 4.1]) can be used to determine whether the optimal controls exist.

Theorem 7. Consider the controlled PHL system

$$\begin{cases} \dot{P}(t) = \Omega - \dfrac{\alpha P(t)H(t)}{N} - (\beta + \mu)P(t) - v(t)P(t), \\ \dot{H}(t) = \dfrac{\alpha P(t)H(t)}{N} - (\gamma + \mu)H(t) - w(t)H(t), \\ \dot{L}(t) = \beta P(t) + \gamma H(t) - \mu L(t) + v(t)P(t) + w(t)H(t), \end{cases}$$

with initial conditions $P(0) \geq 0$, $H(0) \geq 0$, and $L(0) \geq 0$, and the objective functional

$$J(v, w) = H(t_f) - L(t_f) + \int_{t_0}^{t_f} \Big[H(t) - L(t) + \frac{M_1}{2}v^2(t) + \frac{M_2}{2}w^2(t)\Big]dt.$$

Then, there exists an optimal control $(v^*, w^*) \in U_{ad}^2$ such that

$$J(v^*, w^*) = \min_{(v,w)\in U_{ad}^2} J(v, w),$$

provided that all the following conditions hold:

1. The set of corresponding state variables and admissible controls is nonempty.

2. The control set U_{ad} is closed and convex.

3. A linear function in the state and control variables bounds the right-hand side of the state system.

4. The integrand $L(P,H,L,v,w) = H - L + \frac{M_1}{2}v^2 + \frac{M_2}{2}w^2$ of the objective functional is convex on U_{ad}, and there exist constants $c_1, c_2 > 0$ and $\epsilon > 1$ such that

$$L(P,H,L,v,w) \geq -c_1 + c_2\left(|v|^2 + |w|^2\right)^{\epsilon/2}.$$

Proof. **Condition 1.** We must show that for any $(v,w) \in U_{\text{ad}}^2$, the controlled state system admits a unique solution on $[t_0, t_f]$. We use a simplified version of an existing result (see [5, Theorem 7.1.1]).
Let

$$\dot{P} = O_P(t;P,H,L), \quad \dot{H} = O_H(t;P,H,L), \quad \dot{L} = O_L(t;P,H,L),$$

where O_P, O_H, and O_L denote the right-hand sides of the controlled PHL system.

Set the controls $v(t) = c_1$ and $w(t) = c_2$ as constants. Since all parameters are constants, and P, H, and L are continuous, the functions O_P, O_H, and O_L are continuous.
Moreover, the partial derivatives

$$\frac{\partial O_P}{\partial P}, \frac{\partial O_P}{\partial H}, \frac{\partial O_P}{\partial L}, \quad \frac{\partial O_H}{\partial P}, \frac{\partial O_H}{\partial H}, \frac{\partial O_H}{\partial L}, \quad \frac{\partial O_L}{\partial P}, \frac{\partial O_L}{\partial H}, \frac{\partial O_L}{\partial L}$$

are all continuous. Since $N(t) \geq N(0)e^{-\mu t} > 0$ for all finite t (by the positivity established in Section 3), the nonlinear term $\alpha PH/N$ is smooth and the Lipschitz condition is satisfied locally on $\mathbb{R}^3_+$. Consequently, by the Picard–Lindelöf theorem, there exists a unique solution (P,H,L) satisfying the initial conditions. Moreover, by Theorem 1, this solution remains in the bounded invariant set Θ for all $t \in [t_0, t_f]$.
Therefore, the set of state variables and controls is nonempty, and Condition 1 is satisfied.

Condition 2. By the definition, U_{ad} is closed. Let $v_1, v_2 \in U_{ad}$ and $\varepsilon \in [0,1]$; then

$$0 \leq \varepsilon v_1 + (1-\varepsilon)v_2.$$

Since $\varepsilon v_1 \leq \varepsilon$ and $(1-\varepsilon)v_2 \leq 1-\varepsilon$, it follows that

$$\varepsilon v_1 + (1-\varepsilon)v_2 \leq 1.$$

Hence,

$$0 \leq \varepsilon v_1 + (1-\varepsilon)v_2 \leq 1, \quad \forall v_1, v_2 \in U_{ad},\ \varepsilon \in [0,1].$$

A similar argument applies to the control w.

Therefore, U_{ad} is convex and Condition 2 is fulfilled.

Condition 3. Using the differential equations system (1), we consider the total population

$$N(t) = P(t) + H(t) + L(t).$$

Then, we get

$$\frac{dN}{dt} = \Omega - \mu N(t) \leq \Omega - \mu N(t).$$

So,

$$\lim_{t\to\infty} \sup N(t) \leq \frac{\Omega}{\mu}.$$

As a result, every solution for the controlled PHL model is bounded.

Hence, there exist positive constants R_1, R_2, and R_3 such that for all $t \in [t_0, t_f]$,

$$\begin{aligned} P(t) &\leq R_1, \\ H(t) &\leq R_2, \\ L(t) &\leq R_3. \end{aligned}$$

We then consider

$$\begin{cases} O_P = \dot{P}(t) \leq \Omega, \\ O_H = \dot{H}(t) \leq \dfrac{\alpha P(t)H(t)}{N} - w(t)H(t), \\ O_L = \dot{L}(t) \leq \beta P(t) + \gamma H(t) + v(t)R_1 + w(t)R_2. \end{cases}$$

Thus, the system can be rewritten in matrix form as

$$O(t; P, H, L) \leq \bar{A} + BX(t) - RU(t),$$

where

$$O(t; P, H, L) = \begin{bmatrix} O_P \\ O_H \\ O_L \end{bmatrix}, \quad \bar{A} = \begin{bmatrix} \Omega \\ 0 \\ 0 \end{bmatrix}, \quad X(t) = \begin{bmatrix} P \\ H \\ L \end{bmatrix}, \quad U(t) = \begin{bmatrix} v \\ w \end{bmatrix},$$

and

$$B = \begin{bmatrix} 0 & 0 & 0 \\ \frac{\alpha H}{N} & 0 & 0 \\ \beta & \gamma & 0 \end{bmatrix}, \quad R = \begin{bmatrix} 0 & 0 \\ 0 & H \\ -P & -H \end{bmatrix}.$$

The control vector and state variable vector are bounded by a linear function. Consequently, we can write

$$\|O(t; P, H, L)\| \leq \|\bar{A}\| + \|B\|\|X(t)\| + \|R\|\|U(t)\| \leq \psi + \Phi\left(\|X(t)\| + \|U(t)\|\right),$$

where $\psi = \|\bar{A}\|$ and $\Phi = \max(\|B\|, \|R\|)$.

Thus, the sum of the state and control vectors bounds the right-hand side. Consequently, Condition 3 is satisfied.

Condition 4. The integrand in the objective functional

$$J(v, w) = H(t_f) - L(t_f) + \int_{t_0}^{t_f} \left[H(t) - L(t) + \frac{M_1}{2}v^2(t) + \frac{M_2}{2}w^2(t)\right]dt$$

is convex on U_{ad}. The goal is to show that there exist constants $c_1, c_2 > 0$ and $\epsilon > 1$ such that the integrand

$$L(P, H, L, v, w) = H(t) - L(t) + \frac{M_1}{2}v^2(t) + \frac{M_2}{2}w^2(t)$$

satisfies

$$L(P, H, L, v, w) \geq -c_1 + c_2\big(|v|^2 + |w|^2\big)^{\epsilon/2}.$$

Since the state variables $P(t)$, $H(t)$, and $L(t)$ are bounded, we can take

$$\epsilon = 2, \quad c_1 = 2 \sup_{t\in[t_0,t_f]} (H(t), L(t)), \quad c_2 = \inf\left(\frac{M_1}{2}, \frac{M_2}{2}\right).$$

Consequently, it follows that

$$L(P, H, L, v, w) \geq -c_1 + c_2\left(|v|^2 + |w|^2\right).$$

Thus, the integrand of the objective functional is convex and Condition 4 is satisfied. □

In summary, by verifying all four conditions stated in the theorem namely, the non emptiness of the admissible set, the convexity and closedness of the control set, the linear boundedness of the system dynamics, and the convexity of the integrand, it is concluded that an optimal control pair

(v^*, w^*) exists for the controlled PHL model. This result ensures the mathematical soundness of the formulated optimal control problem and provides a rigorous foundation for the derivation of the necessary conditions for optimality, which will be explored in the subsequent subsection.

4.3 Characterization of optimal controls

In this part, we make use of Pontryagin's maximum principle [23]. The key idea is to use the adjoint function to produce the Hamiltonian function by connecting the differential equations system to the objective function. This idea transfers the problem of finding the control to optimize the objective functional subject to the system of differential equations with initial condition and then finds the control to optimize the Hamiltonian pointwise (concerning the control).
Using Pontryagin's maximum principle, the Hamiltonian function is defined as

$$\mathcal{H}(t) = H(t) - L(t) + \frac{M_1}{2}v^2 + \frac{M_2}{2}w^2 + \zeta_1 f_1 + \zeta_2 f_2 + \zeta_3 f_3,$$

where f_1, f_2, f_3 are the right-hand sides of the controlled state system.

Theorem 8. Given an optimal control $v^* = (v^*, w^*) \in U_{ad}^2$ and corresponding solutions P^*, H^*, and L^* of the state system (1), there exist adjoint functions ζ_1, ζ_2, and ζ_3 fulfilling

$$\begin{cases} \dot{\zeta}_1 = \zeta_1\Big\{\alpha\dfrac{H(t)}{N} + \beta + \mu + v(t)\Big\} - \zeta_2\alpha\dfrac{H(t)}{N} - \zeta_3\big\{\beta + v(t)\big\}, \\ \dot{\zeta}_2 = -1 + \zeta_1\alpha\dfrac{P(t)}{N} - \zeta_2\Big\{\alpha\dfrac{P(t)}{N} - \gamma - \mu - w(t)\Big\} - \zeta_3\big\{\gamma + w(t)\big\}, \\ \dot{\zeta}_3 = 1 + \zeta_3\mu \end{cases} \tag{8}$$

At the time t_f, the transversality conditions are given as follows:

$$\begin{aligned} \zeta_1(t_f) &= 0, \\ \zeta_2(t_f) &= 1, \\ \zeta_3(t_f) &= -1. \end{aligned}$$

Moreover, the optimal controls $v^*(t)$ and $w^*(t)$ for $t \in [t_0, t_f]$ are provided by

$$v^*(t) = \min\left\{1, \max\left\{0, \frac{1}{M_1}P(t)(\zeta_1(t) - \zeta_3(t))\right\}\right\}, \tag{9}$$

$$w^*(t) = \min\left\{1, \max\left\{0, \frac{1}{M_2}H(t)(\zeta_3(t) - \zeta_2(t))\right\}\right\}. \tag{10}$$

Proof. The Hamiltonian $\mathcal{H}$ at time t is defined by

$$\begin{aligned}\mathcal{H}(t) = H(t) - L(t) + \frac{M_1}{2}v^2(t) + \frac{M_2}{2}w^2(t)\\ +\zeta_1\Big\{\Omega - \alpha\frac{P(t)H(t)}{N} - (\beta+\mu)P(t) - v(t)P(t)\Big\}\\ +\zeta_2\Big\{\alpha\frac{P(t)H(t)}{N} - (\gamma+\mu)H(t) - w(t)H(t)\Big\}\\ +\zeta_3\Big\{\beta P(t) + \gamma H(t) - \mu L(t) + v(t)P(t) + w(t)H(t)\Big\}.\end{aligned}$$

Using Pontryagin's maximum principle, we obtain the transversality conditions and adjoint equations for $t \in [t_0, t_f]$ as follows:

$$\begin{aligned}\zeta_1(t_f) = 0, \qquad \dot{\zeta}_1(t) = -\frac{\partial\mathcal{H}}{\partial P},\\ \zeta_2(t_f) = 1, \qquad \dot{\zeta}_2(t) = -\frac{\partial\mathcal{H}}{\partial H},\\ \zeta_3(t_f) = -1, \qquad \dot{\zeta}_3(t) = -\frac{\partial\mathcal{H}}{\partial L}.\end{aligned}$$

The optimality condition can be used to solve the optimal controls $v^*(t)$ and $w^*(t)$ for $t \in [t_0, t_f]$:

$$\frac{\partial\mathcal{H}}{\partial v} = M_1 v(t) + \zeta_1(t)(-P(t)) + \zeta_3(t)P(t) = 0,$$

$$\frac{\partial\mathcal{H}}{\partial w} = M_2 w(t) + \zeta_2(t)(-H(t)) + \zeta_3(t)H(t) = 0.$$

That is,

$$\begin{aligned}v(t) = \frac{1}{M_1}P(t)(\zeta_1(t) - \zeta_3(t)),\\ w(t) = \frac{1}{M_2}H(t)(\zeta_3(t) - \zeta_2(t)).\end{aligned}$$

It is straightforward to obtain $v^*(t)$ and $w^*(t)$ in the form of the optimal control laws derived previously, by applying the admissible bounds in U_{ad} for the controls.

Considering the bounds in U_{ad}, the optimal controls are projected as follows:

$$v^*(t) = \min\left\{1, \max\left\{0, \frac{1}{M_1}P(t)(\zeta_1(t) - \zeta_3(t))\right\}\right\},$$

$$w^*(t) = \min\left\{1, \max\left\{0, \frac{1}{M_2}H(t)(\zeta_3(t) - \zeta_2(t))\right\}\right\}.$$

□

In summary, the application of Pontryagin's maximum principle allows us to characterize the necessary conditions for optimal control of the digital platform dynamics. The resulting system, composed of the state equations, the adjoint system, and the optimality conditions, provides a complete framework for determining the optimal strategies $v^*(t)$ and $w^*(t)$. These controls respectively represent the most effective interventions for attracting potential customers and for converting hesitant users into loyal ones, while minimizing associated costs. The obtained results form the theoretical foundation for numerical simulations, which will further illustrate the impact of these optimal strategies on the overall engagement and loyalty within the digital platform ecosystem.

5 Numerical simulations

This subsection is dedicated to the numerical investigation of the controlled PHL model with the aim of evaluating the effectiveness of optimal intervention strategies in influencing customer behavior. Specifically, the model considers two types of customers: Hesitant customers (H), who are undecided about adopting the platform or service, and loyal customers (L), who consistently engage with the platform. By implementing two time-dependent control measures an incentive-based control $v(t)$ and a support-based control $w(t)$ we aim to reduce hesitation, increase loyalty, and optimize the overall performance of the platform within a given time horizon $[t_0, t_f]$.

The numerical simulations are performed using the forward-backward sweep method, which efficiently solves the coupled system of state and adjoint equations derived from Pontryagin's maximum principle. At each iteration, the state system is integrated forward in time using initial conditions, while the adjoint system is integrated backward from the terminal conditions. The controls are updated at each time step according to the optimality conditions, while ensuring that they remain within the admissible range $0 \leq v(t), w(t) \leq 1$. This iterative process continues until convergence is achieved, providing the optimal control trajectories (v^*, w^*) and the corresponding state dynamics. The convergence of this iterative scheme in a finite number of steps is guaranteed under the standard conditions established by Lenhart and Workman [17]: Specifically, when the time horizon T is sufficiently small relative to the Lipschitz constants of the state and adjoint systems, the forward–backward sweep contraction mapping argument ensures that the iterates

converge to the unique optimal solution. In our simulations, convergence is declared when the relative change in the control variables between successive iterations falls below a tolerance of 10^{-6}.

For the numerical simulations, the model parameters were selected based on realistic behavioral assumptions and calibrated to ensure meaningful system dynamics. The baseline parameter values were set as follows: $\Omega = 0.9$, $\alpha = 0.1$, $\beta = 0.4$, $\mu = 0.04$, and $\gamma = 0.01$, with a total population $N = 1400000$. These parameters respectively represent the recruitment rate of potential customers, the interaction rate between potential and hesitant individuals, the spontaneous transition rate toward adoption, the natural decay rate, and the reactivation rate of hesitant individuals. The simulation horizon was fixed at $T = 30$ time units, discretized into $M = 10{,}000$ steps to ensure numerical stability and accuracy.

The initial conditions were defined as $P(0) = 900{,}000$, $H(0) = 500{,}000$, and $L(0) = 0$, representing the initial distributions of potential, hesitant, and loyal customers, respectively. The adjoint variables were initialized with terminal conditions $\zeta_1(T) = 0$, $\zeta_2(T) = 1$, and $\zeta_3(T) = -1$, consistent with the transversality conditions of Pontryagin's maximum principle. The control parameters $M_1 = 60$ and $M_2 = 30$ were introduced as weighting factors in the cost functional, reflecting the relative cost of implementing promotional and awareness strategies. Using these settings, the model was simulated under both uncontrolled and controlled scenarios to analyze the temporal evolution of customer behavior and to evaluate the impact of optimal interventions on e-commerce adoption.

Through these simulations, several scenarios can be explored to understand the impact of control strategies on customer dynamics. For example, we can examine:

- Single-control strategies, where only the incentive-based control $v(t)$ or only the support-based control $w(t)$ is applied.
- Combined control strategies, where both $v(t)$ and $w(t)$ are optimally applied simultaneously.

Additionally, sensitivity analyses of key model parameters, such as the rates of transition between hesitant and loyal customers and the cost weights of the controls, are conducted to identify which factors most strongly influence customer retention and the effectiveness of interventions. These numerical experiments provide a comprehensive understanding of the dynamics of hesitant and loyal customers and offer practical insights into the design of efficient strategies for improving customer loyalty and engagement on digital platforms.

To investigate the evolution of potential, hesitant, and loyal customers and to assess the impact of time-dependent interventions, we implement a numerical scheme for the controlled

PHL model. The procedure involves forward integration of the state system using an explicit Euler method, backward integration of the adjoint system, and iterative updates of the control variables according to Pontryagin's maximum principle. This iterative process continues until convergence of both the state trajectories and control profiles is achieved. The complete numerical resolution algorithm is presented in Algorithm 1.

Algorithm 1 Numerical resolution scheme for the controlled PHL model

Input: Parameters $\Omega, \alpha, \beta, \gamma, \mu, M_1, M_2, N, T, M$
Initialize: $P(0), H(0), L(0)$; $\zeta_1(T), \zeta_2(T), \zeta_3(T)$; $v(0), w(0)$
Set time step $h = T/M$
repeat
    Forward integration (state system)
    for $i = 1$ to M **do**
        Compute derivatives:
        $\dot{P} = \Omega - \alpha P_i H_i/N - (\beta + \mu)P_i$
        $\dot{H} = \alpha P_i H_i/N - (\gamma + \mu)H_i$
        $\dot{L} = \beta P_i + \gamma H_i - \mu L_i$
        Update using Euler scheme:
        $P_{i+1} = P_i + h\dot{P}$
        $H_{i+1} = H_i + h\dot{H}$
        $L_{i+1} = L_i + h\dot{L}$
    end for
    Backward integration (adjoint system)
    for $i = M - 1$ down to 1 **do**
        Compute adjoint derivatives:
        $\dot{\zeta}_1 = \zeta_1(\alpha H_i/N + \beta + \mu + v_i) - \zeta_2(\alpha H_i/N) - \zeta_3(\beta + v_i)$
        $\dot{\zeta}_2 = -1 + \zeta_1(\alpha P_i/N) - \zeta_2(\alpha P_i/N - \gamma - \mu - w_i) - \zeta_3(\gamma + w_i)$
        $\dot{\zeta}_3 = 1 + \zeta_3\mu$
        Update adjoint variables:
        $\zeta_1(i) = \zeta_1(i+1) - h\dot{\zeta}_1$
        $\zeta_2(i) = \zeta_2(i+1) - h\dot{\zeta}_2$
        $\zeta_3(i) = \zeta_3(i+1) - h\dot{\zeta}_3$
    end for
    Update control variables:
    $v_{i+1} = \min(1, \max(0.5, (\zeta_1 - \zeta_3)P_{i+1}/M_1))$
    $w_{i+1} = \min(1, \max(0.5, (\zeta_3 - \zeta_2)H_{i+1}/M_2))$
until Convergence of P, H, L, v, w
Output: Time evolution of $P(t)$, $H(t)$, $L(t)$, $v(t)$, $w(t)$

Interpretation of the Evolution of Hesitant Customers $H(t)$

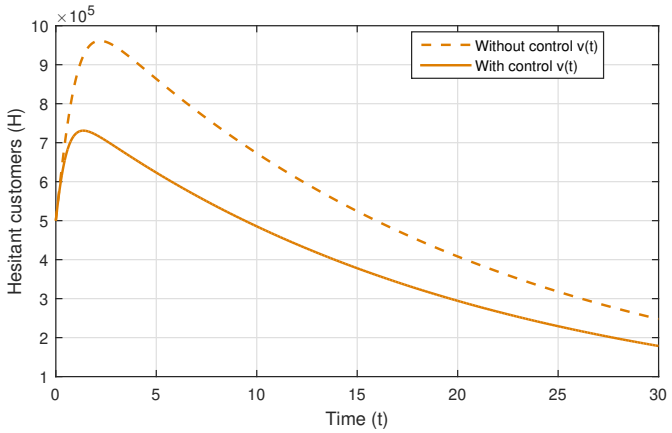


Figure 3: Dynamics of the hesitant model with control v

Figure 3 illustrates the temporal evolution of the hesitant customer population $H(t)$ with and without the implementation of the incentive-based control $v(t)$. In the absence of control, the number of hesitant customers exhibits a sharp and sustained increase, indicating the self-reinforcing nature of rumor propagation on digital platforms. This behavior reflects how misinformation can significantly amplify hesitation among potential users, leading to reduced platform trust and slower conversion to loyalty.

However, when the control $v(t)$ is applied, the curve demonstrates a pronounced decline in the hesitant population over time. This reduction underscores the effectiveness of incentive-driven strategies in mitigating the impact of rumor-induced hesitation. The control essentially acts as a stabilizing mechanism, accelerating the recovery of customer confidence by encouraging engagement and counteracting the deterrent effects of misinformation.

From a practical standpoint, this outcome highlights the role of proactive incentive policies such as reward systems, reputation mechanisms, or verified information campaigns in maintaining user trust. Mathematically, the control $v(t)$ modifies the transition rate from the hesitant to the loyal compartment, thereby reducing the effective reproduction number of hesitation dynamics. The overall effect is a faster convergence toward equilibrium and improved platform stability. In the uncontrolled case, $H(t)$ rises sharply from $H(0) = 500{,}000$, reaching a peak near $t \approx 8$ before declining gradually, a trajectory consistent with the endemic behavior predicted analytically when $\mathcal{R}_0 > 1$. Under control $v(t)$, the peak is both lower in magnitude and reached earlier, with $H(t)$ declining more steeply thereafter, confirming accelerated convergence toward the rumor-free equilibrium $\mathcal{E}_f$, consistent with the global stability result of Theorem 5.

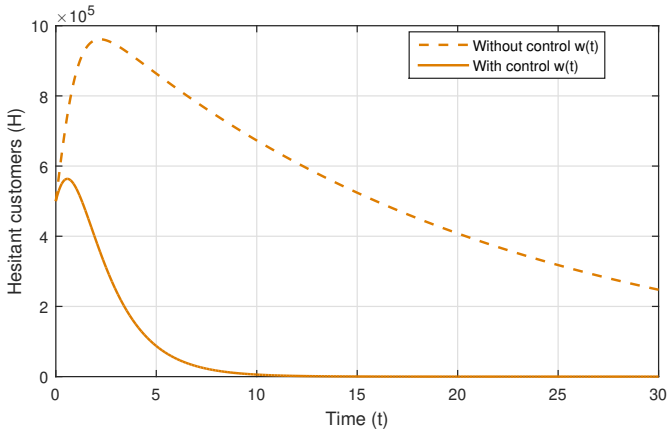


Figure 4: Dynamics of the hesitant model with control w

Figure 4 illustrates the temporal evolution of hesitant customers $H(t)$ under two scenarios: with and without the implementation of the support-based control $w(t)$. The control $w(t)$ represents platform-driven strategies such as user education campaigns, technical support, or trust-building initiatives designed to reduce uncertainty and hesitation among potential users.

In the absence of control, the number of hesitant customers initially rises sharply due to the unmitigated influence of rumor propagation and misinformation within the digital ecosystem. This increase reflects how hesitation naturally amplifies when corrective interventions are lacking, as negative perceptions spread faster among undecided users. Over time, the population of hesitant customers stabilizes at a relatively high level, indicating persistent behavioral stagnation and reduced user conversion toward loyalty.

Conversely, when the control $w(t)$ is applied, a noticeable decline in the hesitant customer population occurs after an initial transient phase. This reduction demonstrates the effectiveness of supportive interventions in mitigating rumor-driven hesitation and encouraging smoother transitions from the hesitant to the loyal customer compartment. The gap between the controlled and uncontrolled trajectories highlights the control's efficiency in fostering platform trust and engagement stability.

Overall, the results confirm that the support-based control $w(t)$ plays a crucial role in counteracting the spread of hesitation by reinforcing user confidence and limiting the long-term effects of misinformation. This finding underscores the importance of sustained and well-calibrated support initiatives in enhancing digital platform resilience and promoting sustained customer loyalty. Compared to the incentive-based control (Figure 3), the support-based control $w(t)$ achieves a faster and deeper suppression of hesitancy in the mid-range of the simulation, since $w(t)$ acts

directly on the H-compartment outflow rate, accelerating the conversion of hesitant users into loyal customers ($\gamma \to \gamma + w(t)$).

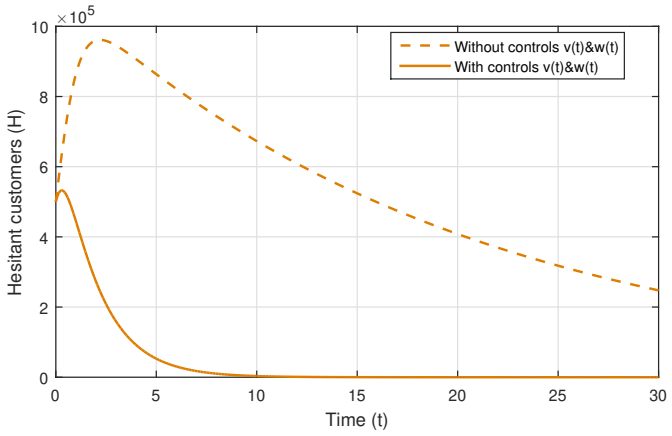


Figure 5: Dynamics of the hesitant model with controls v and w

Figure 5 illustrates the temporal evolution of hesitant customers when both incentive-based $v(t)$ and support-based $w(t)$ interventions are simultaneously implemented. From the plot, we observe that in the absence of controls, the hesitant population exhibits a sustained increase before stabilizing at a relatively high level, indicating that, without strategic intervention, rumor propagation continues to generate uncertainty and hesitation among potential customers. This behavior highlights the self-reinforcing nature of misinformation in digital ecosystems, once hesitation takes hold, it spreads through social influence and weakens customer confidence.

Conversely, under combined control strategies, the curve of hesitant customers declines significantly over time and stabilizes at a much lower equilibrium. This reduction demonstrates the effectiveness of deploying both interventions concurrently: The incentive-based control encourages user engagement and active participation, while the support-based control strengthens trust through guidance and educational measures. The synergy between these two controls accelerates the transition of hesitant users into loyal ones, mitigating the overall impact of rumor-induced hesitation.

Overall, this simulation confirms that multi-faceted intervention strategies outperform isolated approaches, reinforcing the importance of integrated policy design in managing rumor-driven behavioral dynamics. The results align with theoretical predictions from optimal control analysis, validating that the joint optimization of economic incentives and informational support effectively stabilizes customer behavior in the digital environment. This synergistic effect arises because $v(t)$ reduces the inflow of new hesitant users (by converting potential customers directly

to loyalty), while $w(t)$ accelerates the outflow of existing hesitant users toward loyalty, together driving $H(t)$ to near-zero levels by approximately $t \approx 10$, as predicted by Theorems 7 and 8.

Remark 3. The three simulations collectively demonstrate that both incentive-based (v(t)) and support-based (w(t)) controls effectively reduce the number of hesitant customers compared to the uncontrolled case. While each control independently mitigates hesitation, their combined application yields the most significant improvement, accelerating customer transition toward loyalty. This confirms that integrating motivational and informational interventions enhances overall platform stability and resilience against rumor-induced hesitation.

Interpretation of the Evolution of Loyal Customers $L(t)$

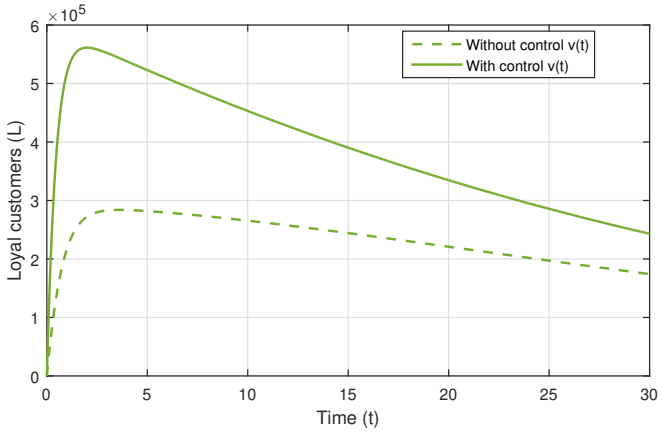


Figure 6: Dynamics of the loyal model with control v

Figure 6 illustrates the temporal evolution of the loyal customer population $L(t)$ under two scenarios: With and without the incentive-based control $v(t)$. In the absence of control, the number of loyal customers exhibits a slow and modest increase, stabilizing around approximately 30,000 individuals by the end of the simulation period. This limited growth reflects the natural rate of customer transition toward loyalty when no proactive interventions are applied, leaving the system more vulnerable to rumor-driven hesitation.

In contrast, when the control $v(t)$ is implemented, the curve shows a rapid and substantial rise in loyalty, reaching nearly 85,000 loyal customers over the same time horizon. This pronounced improvement highlights the effectiveness of incentive mechanisms such as promotional campaigns, loyalty rewards, or positive information reinforcement in converting hesitant users into committed customers. Moreover, the controlled trajectory remains stable throughout, indicating that the strategy effectively mitigates the destabilizing effects of rumor dissemination.

Overall, these numerical outcomes demonstrate that the application of $v(t)$ not only enhances customer retention but also fortifies the system's resilience against misinformation, ensuring a more robust and trustworthy digital platform ecosystem.

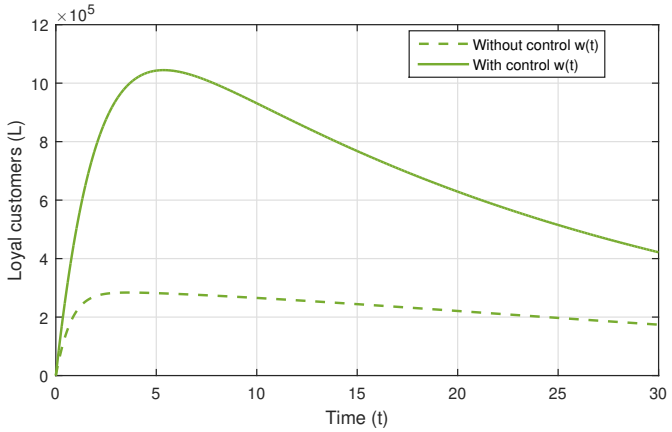


Figure 7: Dynamics of the loyal model with control w

The results illustrated in Figure 7 compare the evolution of loyal customers $L(t)$ under two scenarios: With and without the implementation of the control $w(t)$. Without control, the loyal customer population remains relatively stagnant, stabilizing near 6.0×10^5 by the end of the simulation period. In contrast, when the control is activated, $L(t)$ exhibits a marked increase, reaching approximately 1.2×10^6 after 30 time units, almost a twofold improvement.

This outcome underscores the crucial role of $w(t)$ in mitigating the negative influence of rumor propagation on user confidence and loyalty. By reducing rumor-induced hesitation, the control promotes a faster and more sustained transition of users toward loyalty. The observed acceleration in loyalty growth highlights the efficiency of targeted rumor refutation and trust reinforcement interventions, confirming the effectiveness of the proposed strategy in enhancing both user engagement and platform stability. The improvement is sustained and monotone from the early stages of the simulation, reflecting the direct impact of $w(t)$ on the $H \to L$ transition rate, and has direct practical relevance for platform managers: Investment in user support infrastructure (e.g., FAQ systems, live assistance, trust verification badges) produces long-run loyalty gains that outweigh short-term implementation costs.

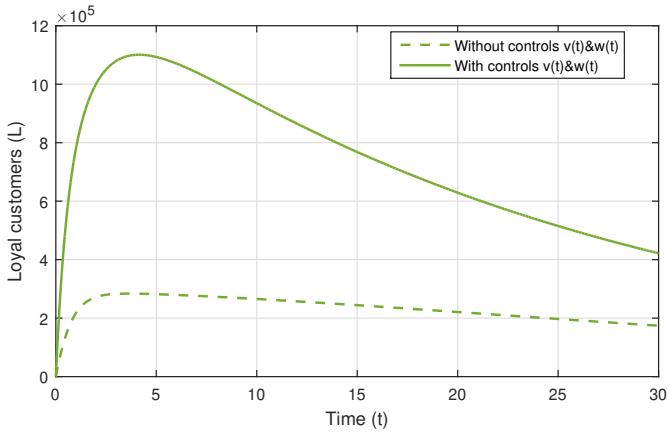


Figure 8: Dynamics of the loyal model with controls v and w

The simulation depicted in Figure 8 illustrates the temporal evolution of loyal customers $L(t)$ under both controlled and uncontrolled scenarios involving the two strategies $v(t)$ and $w(t)$. Initially, the population of loyal users increases gradually. However, in the absence of controls, the curve stabilizes at a significantly lower level around 6×10^5, reflecting the persistent impact of rumor propagation, which erodes trust and discourages long-term engagement.

When both control strategies are simultaneously applied, the growth of loyal customers accelerates markedly, reaching nearly 1.2×10^6 by the end of the simulation period ($t = 30$). This represents an approximate doubling in loyalty rates compared to the uncontrolled case. The combined influence of $v(t)$ (incentive-based) and $w(t)$ (support-based) interventions effectively counteracts the hesitation induced by rumors, restoring confidence and fostering retention.

Overall, the figure underscores the synergistic effect of applying integrated control mechanisms: While rumor dynamics naturally hinder user loyalty, coordinated actions to both incentivize potential users and reinforce hesitant ones can significantly mitigate misinformation's adverse impact, leading to a more resilient and trustworthy digital system.

Remark 4. The comparative analysis of the three figures highlights the effectiveness of the proposed control strategies in mitigating rumor-induced hesitation and enhancing customer loyalty. While uncontrolled dynamics reveal the detrimental influence of rumor dissemination on customer engagement, the implementation of the incentive based control $v(t)$ and the support based control $w(t)$, especially when applied jointly demonstrates a substantial improvement in stabilizing hesitant users and promoting loyalty. These findings confirm that coordinated interventions not only suppress the negative behavioral impact of rumors but also foster sustainable platform growth.

6 Conclusion

This study proposed and analyzed an optimal control model that captures the rumor-induced behavioral dynamics of customers on digital platforms. Drawing inspiration from epidemiological modeling, the developed PHL framework divides the customer population into potential, hesitant, and loyal compartments, with transitions driven by the diffusion of rumors. The model formalizes how misinformation can propagate through user interactions, fostering hesitation and reducing loyalty, while also providing a means to assess and mitigate these effects through optimal control strategies. The incorporation of incentive-based and support-based control functions enables the design of interventions that not only suppress the negative impact of rumors but also encourage customer engagement and trust restoration. Theoretical analysis, supported by Pontryagin's maximum principle, established the existence and characterization of optimal controls, offering a systematic foundation for decision-making. Sensitivity analysis of the basic reproduction number (R_0) revealed the dominant parameters influencing rumor persistence, highlighting which behavioral or informational levers are most effective in minimizing the spread of hesitation. Numerical simulations conducted using a forward–backward sweep method based on an explicit Euler discretization scheme demonstrated the practical implications of these findings. The results showed that the combined application of incentive and support controls markedly reduces the hesitant population, curtails rumor propagation, and accelerates the growth of loyal users, thereby reinforcing platform stability and credibility. In summary, this research bridges rumor dynamics, behavioral modeling, and optimal control theory into a unified mathematical approach for digital ecosystems. It provides both analytical depth and strategic insight into how misinformation can be contained and counteracted. Future studies may extend this model to stochastic or network-based settings, integrate real-time data, or explore adaptive rumor countermeasures, further enhancing the applicability of mathematical control frameworks in addressing the societal and economic challenges posed by digital misinformation.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Amini, A., Bayiz, Y.E., Lee, E.-J., Somer-Topcu, Z., Mărculescu, R. and Topcu, U. *How media competition fuels the spread of misinformation*, Sci. Adv. 11(25) (2025), eadu7743..

[2] Baek, J., Dinev, A. and Lykouris, T. *Social learning with bounded rationality: Negative reviews persist under newest first*, In Proceedings of the 25th ACM Conference on Economics and Computation (pp. 375–375) 2024.

[3] Bakenova, K., Kuznetsov, A., Artyshchuk, I., Shaikhanova, A., Shevchuk, R. and Orobchuk, O. *Information diffusion modeling in social networks: A comparative analysis of delay mechanisms using population dynamics*, Appl. Sci. 15(11) (2025), 6092.

[4] Belhdid, S., Balatif, O. and Khajji, B. *Optimal control and cost-effectiveness analysis of scam rumor propagation over social networks*, Results Control Optim. 16 (2024), 100441.

[5] Boyce, W.E. and DiPrima, R.C. *Elementary differential equations and boundary value problems*, John Wiley and Sons, New York, 2009.

[6] Chen, J., Zhang, L. and Wang, X. *SEIOR rumor propagation model considering hesitating mechanism*, Mathematics 11(2) (2023), 283.

[7] Danenberg, T. and Fudenberg, D. *Endogenous attention and the spread of false news*, arXiv:2406.11024 (2024).

[8] Ding, J., Li, Z., Wu, X., Liu, R. and Hu, H. *Information dissemination model based on social networks characteristics*, Mathematics 13(8) (2025), 1254.

[9] El Bhih, A., Yaagoub, Z., Rachik, M. and Allali, K. *Controlling the dissemination of rumors and antirumors in social networks: A mathematical modeling and analysis approach*, Eur. Phys. J. Plus 139(2) (2024), 1–23.

[10] Fleming, W.H. and Rishel, R.W. *Deterministic and stochastic optimal control*, Springer, New York, 1975.

[11] Gonen, L.D., Tavor, T. and Spiegel, U. *Unlocking market potential: Strategic consumer segmentation and dynamic pricing for balancing loyalty and deal seeking*, Mathematics 12(21) (2024), 3364.

[12] Iranmanesh, A. and Rahgozar, M. *IMEGAR: An enhanced hybrid genetic algorithm for dynamic influence maximization in social networks*, Research Square (2024).

[13] Jakobsen, L., Holden, A.J., Gürcan, Ö. and Özgöbek, Ö. *Agent-based exploration of recommendation systems in misinformation propagation*, In Social Simulation Conference, pp. 269-284. Cham: Springer Nature Switzerland, 2025.

[14] Kang, S., Hou, X., Hu, Y. and Liu, H. *Dynamic analysis and optimal control considering cross transmission and variation of information*, Sci. Rep. 12(1) (2022), 17245.

[15] Laaroussi, A.E.A., El Bhih, A. and Rachik, M. *Optimal vaccination and treatment policies with constrained inequalities to study limited vaccination resources for a multistrain reaction-diffusion SEIR model of COVID-19*, Partial Differ. Equ. Appl. Math. 10 (2024), 100684.

[16] LaSalle, J.P. *The stability of dynamical systems*, Regional Conference Series in Applied Mathematics, Vol. 25, SIAM, Philadelphia, 1976.

[17] Lenhart, S. and Workman, J.T. *Optimal control applied to biological models*, Chapman & Hall/CRC, Boca Raton, 2007.

[18] Li, D., Zhao, Y. and Deng, Y. *Rumor spreading model with a focus on educational impact and optimal control*, Nonlinear Dyn. 112(2) (2023), 1575–1592.

[19] Li, S. and Zhang, Q. *Dynamics analysis and optimal control of rumor propagation in online platforms*, Comput. Math. Methods Med. (2025), 5503137.

[20] Liu, F., Zhang, Z. and Buss, M. *Robust optimal control of deterministic information epidemics with noisy transition rates*, Physica A 517 (2018), 577–589.

[21] López, A.B., Pastor-Galindo, J. and Ruipérez-Valiente, J.A. *Frameworks, modeling and simulations of misinformation and disinformation: A systematic literature review*, arXiv:2406.09343 (2024).

[22] Mansal, F. and Faye, I. *Mathematical modeling and optimal control of untrue information: Dynamic SEIZ in online social networks*, arXiv:2309.13058 (2023).

[23] Pontryagin, L.S., Boltyanskii, V.G., Gamkrelidze, R.V. and Mishchenko, E.F. *The mathematical theory of optimal processes*, Wiley, New York, 1962.

[24] Van den Driessche, P. and Watmough, J. *Reproduction numbers and subthreshold endemic equilibria for compartmental models of disease transmission*, Math. Biosci. 180(1–2) (2002), 29–48.

[25] Wu, H. and Li, J. *Analysis and control of rumor propagation model with repeated waiting behavior*, Mathematics 13(2) (2025), 312.

[26] Yaagoub, Z., El Bhih, A. and Allali, K. *Global analysis of a fractional-order infection model for the propagation of computer viruses*, Model. Earth Syst. Environ. 11(1) (2025), 68.

[27] Zheng, W. and Liu, Y. *Research on rumor and anti-rumor propagation models in social networks*, Sci. Rep. 15 (2025), 96006.

Iranian Journal of Numerical Analysis and Optimization
Vol. 17, No. 1, 2027, pp 109–144
https://doi.org/10.22067/ijnao.2026.98939.1878
https://ijnao.um.ac.ir/

CC BY NC

Research Article

Understanding the impact of vaccination on Omicron variants

G.P. Sahu, A.S. Thakur*, C.S. Sikarwar and H. Singh

Abstract

The COVID-19 pandemic, caused by the SARS-CoV-2 virus, has led to the emergence of multiple variants, including Alpha, Beta, Gamma, Delta, and Omicron, each exhibiting distinct transmissibility and severity profiles. In this context, a new deterministic mathematical model has been developed as an extension of Thakur and Sahu [51], designed to capture the dynamics of COVID-19 transmission while incorporating the effects of both the first and second doses of vaccination. The model is examined using the theory of dynamical systems. The basic

*Corresponding author

Received 26 May 2026; revised 24 July 2026; accepted 5 August 2026

Govind Prasad Sahu
Center for Basic Sciences, Pt. Ravishankar Shukla University, Raipur, Chhattisgarh. e-mail: govind3012@gmail.com

Amit Singh Thakur
School of Studies in Mathematics, Pt. Ravishankar Shukla University, Raipur, Chhattisgarh. e-mail: amitsinghprsu@gmail.com

Chhatrpal Singh Sikarwar
Department of Mathematics, Govt Kachna Dhurwa College Chhura, Chhattisgarh. e-mail: chhatrapal84@gmail.com

Harkaran Singh
College of Trades, Technology, and Hospitality, Victoria University, Footscray, Victoria 3011, Australia. e-mail: harkaran1978@gmail.com

How to cite this article
Sahu, G.P., Thakur, A.S., Sikarwar, C.S. and Singh, H., Understanding the impact of vaccination on Omicron variants. *Iran. J. Numer. Anal. Optim.*, 2027; 17(1): 109-144. https://doi.org/10.22067/ijnao.2026.98939.1878

reproduction number is derived, and both the local stability of the disease-free and endemic equilibria, as well as the global stability of the disease-free equilibrium, are established. The model is calibrated using data from India's third wave of the COVID-19 pandemic. Sensitivity analysis, employing partial rank correlation coefficients and normalized forward sensitivity indices, identifies vaccination rates as key parameters driving disease dynamics. Findings reveal that the Omicron variant has the potential to cause large-scale outbreaks even with ongoing vaccination campaigns. The study emphasizes the need for sustained high vaccination coverage and the continued implementation of nonpharmaceutical interventions to control the pandemic effectively. Furthermore, the model is extended to an optimal control framework, concluding that dynamically adjusting five targeted control measures offers the most effective strategy for reducing infections while minimizing associated costs.

AMS subject classifications (2020): Primary 92D30; Secondary 92C60.

Keywords: COVID-19; Environmental transmission; Vaccination; Optimal control; Omicron.

1 Introduction

COVID-19, caused by the SARS-CoV-2 virus, has emerged as one of the most significant public health crises of the 21st century, profoundly impacting global health, economies, and societies. Since its emergence in late 2019, the virus has spread rapidly worldwide, exposing vulnerabilities in healthcare infrastructure and pandemic preparedness [57]. During the early stages of the pandemic, when biomedical interventions were not yet widely available, nonpharmaceutical measures—such as spreading information through television, radio, social media, and the Internet—played a crucial role in reducing disease transmission [50, 49]. Media coverage played a critical role in influencing public behaviour, encouraging preventive actions such as mask-wearing, social distancing, quarantine, and isolation [35]. While these measures reduced transmission risk, they were insufficient on their own to contain the pandemic [15].

Vaccination quickly became central to mitigating the impact of COVID-19, as vaccines prepared the immune system to fight infection [45, 29]. By early 2021, the distribution of vaccines had significantly improved the global pandemic outlook. In India, the vaccination campaign began on January 16, 2021, and by December 31, 2021, approximately 89% of the population had received at least one dose [48]. Despite these achievements, SARS-CoV-2 continued to evolve, with variants such as Alpha, Beta, Gamma, Delta, and Omicron emerging [10]. Although vaccines developed against early variants exhibited reduced efficacy against Delta and Omicron, they continued to provide substantial protection against severe disease [23, 47]. The Delta variant

in particular demonstrated increased transmissibility, shorter incubation periods, higher viral loads, and greater disease severity [9].

The Omicron variant, first reported in November 2021 in South Africa [36], rapidly became dominant worldwide. In India, Omicron's prevalence rose dramatically, from 0.35% to over 97% between November 29, 2021, and February 7, 2022 [44]. This surge led to record infection rates in cities such as Mumbai, where daily cases escalated from 191 to over 20,000 by early January 2022 [44]. Omicron exhibited higher transmissibility than previous variants, contributing to widespread infections, particularly among younger, less-vaccinated populations [31, 18].

Mathematical modeling has played a pivotal role in understanding COVID-19 transmission dynamics and guiding public health interventions, especially in India [3, 24, 41, 5, 27, 6, 40]. Researchers have developed various compartmental models to simulate the spread of disease and assess control strategies. For instance, Bagal et al. [3] simulated COVID-19 transmission, Bajiya et al. [5] emphasized the role of face masks and quarantine, and Negi et al. [27] introduced an SEIAHR framework incorporating vaccination and quarantine. Bandekar et al. [6] analyzed India's third wave, highlighting the benefits of mask compliance, screening, and timely treatment. Sahu and Thakur [38] analyzed how media influence, environmental factors, and treatment capacity constraints affect the spread of epidemics in India.

Recent modeling studies have recognized the importance of environmental contamination in transmission [40, 26, 51, 37]. The virus can remain viable on surfaces for days, and infected individuals contribute to environmental contamination through respiratory droplets [14, 54]. Modern epidemic models have evolved beyond classical SEIR frameworks [32], incorporating deaths [42, 43], isolation [34], asymptomatic infection [39], hospitalization [41], and vaccination strategies [27, 22, 46, 11], thus enabling realistic simulation of interventions.

Several studies have specifically assessed vaccination's role in pandemic control, including multi-dose strategies [12, 27, 22, 46, 11, 4, 17, 56]. By the end of 2021, 64% of India's population was fully vaccinated with two doses [48], and two doses of Covishield or Covaxin significantly reduced infection risk according to the Indian Council of Medical Research.

In this work, we develop a mathematical modeling framework to investigate the combined effects of vaccination and environmental transmission on the spread of COVID-19 in India during the Omicron wave. The proposed model incorporates partially and fully vaccinated populations together with environmental viral dynamics [17, 56, 51], enabling a realistic assessment of intervention strategies. Using real epidemiological data and optimal control techniques, the study aims to guide policymakers in designing effective mitigation measures against future SARS-CoV-2 outbreaks [1, 5, 28, 58, 19].

The organization of this work is as follows: Section 2 presents a novel mathematical model that captures the dynamics of COVID-19 transmission. In Section 3, we examine the model's

well-posedness, characterize equilibrium points, analyze stability properties, and derive the control reproduction number along with possible bifurcations. Section 4 focuses on numerical analyses, including parameter estimation from real-world data, sensitivity assessments using normalized and partial rank correlation coefficients (PRCC) methods, and simulations evaluating the influence of dual vaccine doses and environmental factors on disease spread. Section 5 formulates and solves an optimal control problem related to the model. Finally, Section 6 summarizes the main findings and outlines future research directions.

2 Mathematical model

The human population $N(t)$ within a given community is divided into eight compartments, each representing a different stage in the disease dynamics: susceptible individuals $S(t)$, exposed individuals $E(t)$, asymptomatic carriers $I_a(t)$, symptomatic infected individuals $I(t)$, those infected with the Omicron variant $I_0(t)$, individuals who have received the first vaccine dose $V_1(t)$, those who have received the second dose $V_2(t)$, and recovered individuals $R(t)$. Additionally, the environmental concentration of the virus is denoted by $U(t)$. Hence, the total population at any time t is given by

$$N(t) = S(t) + E(t) + I_a(t) + I(t) + I_0(t) + V_1(t) + V_2(t) + R(t).$$

We propose the following assumptions:

- The population is uniformly spread out, with new individuals consistently entering the susceptible group at a fixed rate Λ, while natural deaths occur at a constant rate μ.
- The parameter β represents the transmission rate at which a susceptible individual contracts an infection through valid contact with infected individuals (asymptomatic, symptomatic, and Omicron-infected individuals). Parameters η_1 and η_2 represent the relative infectiousness of individuals with symptomatic and Omicron variant infections, respectively, compared to asymptomatic individuals.
- The latent period of infected individuals is represented by $1/\sigma$. After the latent period, exposed individuals who do not exhibit clinical symptoms join the asymptomatic infected class at a rate of $\sigma\psi$. Individuals who develop clinical symptoms of COVID-19 progress to the symptomatic infected compartment at a rate $\sigma(1-\psi-\phi)$. In contrast, the remaining exposed individuals transition to the Omicron-infected compartment at a rate $\phi\sigma$.

- Individuals at risk of COVID-19 receive a first vaccine dose (Dose I) at rate ν_1, providing partial immunity. However, this immunity wanes over time, and individuals vaccinated with Dose I gradually return to the susceptible class at rate ω_1 [14, 11]. Additionally, protection from Dose I is incomplete, leaving some individuals partially susceptible, with reduced susceptibility quantified by the factor θ [56]. Administration of a second vaccine dose (Dose II) at rate ν_2 leads to optimal protection, substantially lowering the risk of reinfection [4, 17, 21].

- Symptomatic individuals transition to the recovered category at a recovery rate of γ_2, whereas those infected with the Omicron variant recover at a different rate, γ_3. Asymptomatic cases, which do not show visible symptoms, recover on their own and join the recovered group at a rate γ_1. Over time, immunity in recovered individuals wanes, causing a portion of them to revert to the susceptible group at a rate denoted by χ.

- Individuals showing symptoms and those infected with the Omicron variant succumb to the disease at respective rates denoted by δ_1 and δ_2.

- Individuals infected with COVID-19 emit the virus into the environment primarily via respiratory activities such as coughing and sneezing. The environmental contribution rates of asymptomatic, symptomatic, and Omicron-infected individuals are represented by θ_1, θ_2, and θ_3, respectively. Despite this shedding, the virus has a limited environmental lifespan due to factors such as disinfection efforts and natural decay, with its removal from the environment occurring at a rate ϵ.

Based on the above assumptions, the spread of COVID-19 can be modeled using the following system of nonlinear ordinary differential equations:

$$
\begin{cases}
\frac{dS}{dt} &= \Lambda - \frac{\beta(\eta_1 I + I_a + \eta_2 I_0)S}{N} - \beta_1 U S - \nu_1 S + \omega_1 V_1 + \chi R - \mu S, \\
\frac{dE}{dt} &= \frac{\beta(\eta_1 I + I_a + \eta_2 I_0)(S + \theta V_1)}{N} + \beta_1 U S - \sigma E - \mu E, \\
\frac{dI_a}{dt} &= \psi\sigma E - \gamma_1 I_a - \mu I_a, \\
\frac{dI}{dt} &= (1 - \psi - \phi)\,\sigma E - \gamma_2 I - \delta_1 I - \mu I, \\
\frac{dI_0}{dt} &= \phi\sigma E - \gamma_3 I_0 - \delta_2 I_0 - \mu I_0, \\
\frac{dV_1}{dt} &= \nu_1 S - \nu_2 V_1 - \omega_1 V_1 - \frac{\beta(\eta_1 I + I_a + \eta_2 I_0)\theta V_1}{N} - \mu V_1, \\
\frac{dV_2}{dt} &= \nu_2 V_1 - \gamma_4 V_2 - \mu V_2, \\
\frac{dR}{dt} &= \gamma_2 I + \gamma_1 I_a + \gamma_3 I_0 + \gamma_4 V_2 - \chi R - \mu R, \\
\frac{dU}{dt} &= \theta_2 I + \theta_1 I_a + \theta_3 I_0 - \epsilon U.
\end{cases}
\tag{1}
$$

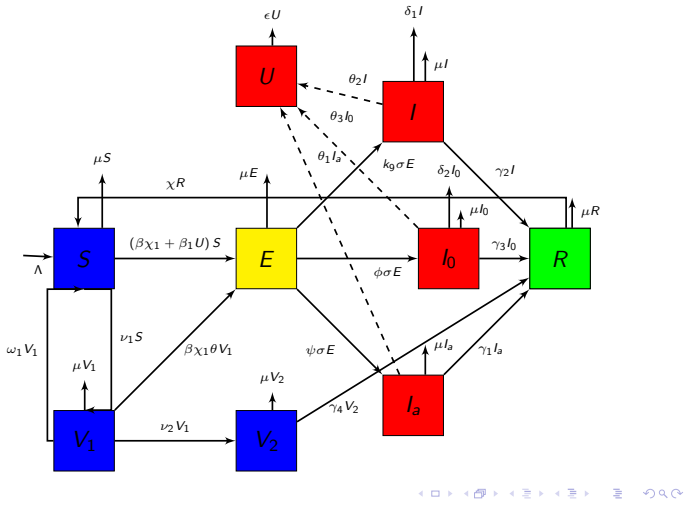


Figure 1: Flow diagram

A schematic representation of the model described by (1) is illustrated in Figure 1. The model begins with the following initial conditions:

$$S(0) = S_0 > 0, \quad E(0) \geq 0, \quad I_a(0) \geq 0, \quad I(0) \geq 0, \quad I_0(0) \geq 0,$$

$$V_1(0) = V_{10} > 0, \quad V_2(0) = V_{20} > 0, \quad R(0) \geq 0, \quad U(0) \geq 0.$$

To simplify notation, we define the following constants:

$$\begin{aligned}
&k_1 = \nu_1 + \mu, && k_2 = \mu + \sigma, && k_3 = \gamma_1 + \mu,\\
&k_4 = \gamma_2 + \delta_1 + \mu, && k_5 = \gamma_3 + \delta_2 + \mu, && k_6 = \nu_2 + \omega_1 + \mu,\\
&k_7 = \gamma_4 + \mu, && k_8 = \chi + \mu, && k_9 = 1 - \phi - \psi,\\
&\chi_1 = \frac{\eta_1 I + I_a + \eta_2 I_0}{N}.
\end{aligned}$$

Accordingly, the system governed by (1) can be rewritten as follows:

$$\begin{cases} \frac{dS}{dt} &= \Lambda - \frac{\beta(\eta_1 I + I_a + \eta_2 I_0)S}{N} - \beta_1 US + \omega_1 V_1 + \chi R - k_1 S, \\ \frac{dE}{dt} &= \frac{\beta(\eta_1 I + I_a + \eta_2 I_0)(S+\theta V_1)}{N} + \beta_1 US - k_2 E, \\ \frac{dI_a}{dt} &= \psi\sigma E - k_3 I_a, \\ \frac{dI}{dt} &= k_9\sigma E - k_4 I, \\ \frac{dI_0}{dt} &= \phi\sigma E - k_5 I_0, \\ \frac{dV_1}{dt} &= \nu_1 S - \frac{\beta(\eta_1 I + I_a + \eta_2 I_0)\theta V_1}{N} - k_6 V_1, \\ \frac{dV_2}{dt} &= \nu_2 V_1 - k_7 V_2, \\ \frac{dR}{dt} &= \gamma_2 I + \gamma_1 I_a + \gamma_3 I_0 + \gamma_4 V_2 - k_8 R, \\ \frac{dU}{dt} &= \theta_2 I + \theta_1 I_a + \theta_3 I_0 - \epsilon U. \end{cases} \tag{2}$$

Table 1: Parameters and their meanings in relation to system (1)

Parameters	Description
Λ	Recruitment rate
$1/\chi$	Average duration of immunity before waning
θ_1	Viral shedding rate by asymptomatic individuals
θ_2	Viral shedding rate by symptomatic individuals
θ_3	Viral shedding rate by Omicron-infected individuals
ω_1	Rate of loss of vaccine-induced immunity
η_1	Adjustment parameter
η_2	Adjustment parameter
$1/\sigma$	Latent period
ψ	Modification parameter
ϕ	Fraction of exposed individuals showing symptoms
ϵ	Decline rate of viral viability over time in contaminated surroundings
δ_1	Mortality rate due to disease in non-Omicron-infected individuals
δ_2	Disease-induced death rate of Omicron-infected individuals
γ_1	Recovery rate for asymptomatic individuals
μ	Natural mortality rate
β	The rate at which the infection spreads
β_1	The highest rate of spread caused by environmental contamination
ν_1	Vaccination rate (dose I)
ν_2	Vaccination rate (dose II)
γ_2	Rate of recovery in those exhibiting symptoms
γ_3	Recovery rate for Omicron-infected individuals
γ_4	Rate of progression to immune class as a result of receiving the second dose of the vaccine
θ	Relative risk of infection for vaccinated individuals with the dose I

Table 2: Model (1) calibration values

Parameters	Range	Baseline	Sources
β	$(0.2, 0.8999)\ day^{-1}$	0.6594	[52]
β_1	$(0, 0.000105)\ day^{-1}$	0.000104	[40]
ν_1	$(0, 1)\ day^{-1}$	0.0215	[55]
ν_2	$(0, 1)\ day^{-1}$	0.0341	[55]
γ_1	$(0.1, 0.5)\ day^{-1}$	0.3899	Estimated
γ_2	$(0.001, 0.2)\ day^{-1}$	0.069	Estimated
γ_3	$(0.01, 0.3)\ day^{-1}$	0.1746	Estimated
γ_4	$(0.1, 0.9)\ day^{-1}$	0.84	[17]
δ_1	$(0, 0.1)\ day^{-1}$	0.08	Estimated
δ_2	$(0, 0.09)\ day^{-1}$	0.0015	Estimated
ψ	$(0.9, 0.98)\ day^{-1}$	0.9566	[20]
ϕ	$(0.01, 0.02)\ day^{-1}$	0.0101	[20]
Λ	$(10, 50000)\ day^{-1}$	1319.294	[1]
μ	$(0, 0.00042)\ day^{-1}$	0.000042578	[1]
θ_1	$(0, 0.0000021)\ day^{-1}$	0.0000012	[40]
θ_2	$(0, 0.00013)\ day^{-1}$	0.00011	[40]
θ_3	$(0, 0.001)\ day^{-1}$	0.0001	Assumed
ϵ	$(0, 0.1)\ day^{-1}$	0.09	[40]
σ	$(0.2, 0.9)\ day^{-1}$	0.8999	[20]
ω_1	$(0, 0.61)\ day^{-1}$	0.1224	[17]
χ		$0.0027\ day^{-1}$	Assumed
η_1	$(0.45, 0.9)$	0.78	[20, 13]
η_2	$(0.35, 0.77)$	0.49	[20, 13]
θ	$(0.01, 0.5)\ day^{-1}$	0.315	[56]

3 Mathematical analysis

In this section, we present key analytical insights for the model (2), including the positivity and boundedness of its solutions, the existence of both disease-free and endemic equilibria, and an analysis of their stability with respect to the control reproduction number.

3.1 Nonnegativity and boundedness

This section shows that the solutions to the model system (1) remain both nonnegative and bounded.

$\frac{dS}{dt}\big|_{S=0} = \Lambda + \omega_1 V_1 + \chi R > 0,$ $\quad \frac{dE}{dt}\big|_{E=0} = \frac{\beta(\eta_1 I + I_a + \eta_2 I_0)S}{N} + \beta_1 US \geq 0,$

$\frac{dI_a}{dt}\big|_{I_a=0} = \psi\sigma E \geq 0,$ $\quad \frac{dI}{dt}\big|_{I=0} = k_9\sigma E \geq 0,$ $\quad \frac{dI_0}{dt}\big|_{I_0=0} = \phi\sigma E \geq 0,$

$\frac{dV_1}{dt}\big|_{V_1=0} = \nu_1 S > 0,$ $\quad \frac{dV_2}{dt}\big|_{V_2=0} = \nu_2 V_1 > 0,$

$\frac{dR}{dt}\big|_{R=0} = \gamma_2 I + \gamma_1 I_a + \gamma_3 I_0 + \gamma_4 V_2 > 0, \quad \frac{dU}{dt}\big|_{U=0} = \theta_2 I + \theta_1 I_a + \theta_3 I_0 \geq 0.$

Since all the rates in the system are nonnegative, starting within the interior of the nonnegative cone R^9_+ ensures the system's trajectory remains inside it. This is because the vector field points inward along every boundary surface, guaranteeing that none of the solutions to system (1) will become negative.

To demonstrate that the solutions are bounded, we consider the full system (1), leading to the equation:

$$\frac{dN}{dt} = \Lambda - \mu N - \delta_1 I - \delta_2 I_0.$$

Since $\delta_1 I + \delta_2 I_0 \geq 0$, it follows that

$$\frac{dN}{dt} < \Lambda - \mu N.$$

From [7], we conclude as $t \to \infty$ that

$$0 \leq N(t) \leq \frac{\Lambda}{\mu} = N_0.$$

Because $I(t), I_0, I_a \leq N(t)$ at any time, we also have

$$I(t), I_0, I_a \leq \frac{\Lambda}{\mu}.$$

Next, from (1), we derive

$$\frac{dU}{dt} = \theta_2 I + \theta_1 I_a + \theta_3 I_0 - \epsilon U \leq (\theta_2 + \theta_1 + \theta_3)\frac{\Lambda}{\mu} - \epsilon U.$$

Since $0 \leq U(0)$, applying Gronwall's inequality yields

$$U(t) \leq \frac{(\theta_2 + \theta_1 + \theta_3)\Lambda}{\epsilon\mu}.$$

Hence the feasible region for the model system (1) is

$$\Theta = \left\{ (S, E, I_a, I, I_0, V_1, V_2, R, U) \in R^9_+ : \right.$$
$$0 \leq S, E, I_a, I, I_0, V_1, V_2, R, N \leq \frac{\Lambda}{\mu},$$
$$\left. 0 \leq U \leq \frac{(\theta_2 + \theta_1 + \theta_3)\,\Lambda}{\epsilon\mu} \right\}.$$

The preceding analysis of nonnegativity and boundedness leads to the following conclusion.

Theorem 1. Under the given initial conditions, the system's solutions remain nonnegative for all future times. Moreover, this closed region is positively invariant under the flow of the model system (1).

3.2 Disease-free equilibrium and control reproduction number

To determine the model's control reproduction number R_c, we apply the next-generation matrix method as outlined in [53]. This yields

$$\begin{aligned}\mathcal{R}_c =& \frac{1}{k_2k_3k_4k_5(r_0+S_0+V_{10}+V_{20})\epsilon}\sigma(\beta\epsilon(S_0+V_{10}\theta)c_1\\ &+S_0(r_0+S_0+V_{10}+V_{20})\beta_1c_2),\end{aligned}$$

where

$$c_1 = (k_3k_5k_9\eta_1 + k_4k_3\eta_2\phi + k_4k_5\psi),$$

$$c_2 = (k_3k_5k_9\theta_2 + k_4k_3\theta_3\phi + k_4k_5\theta_1\psi),$$

$$c_3 = (k_3k_5k_9\gamma_2 + k_4k_3\gamma_3\phi + k_4k_5\gamma_1\psi),$$

$$\phi = \frac{\beta(S_0+\theta V_{10})}{r_0+S_0+V_{10}+V_{20}}.$$

Additionally, the model (2) admits the disease-free equilibrium (DFE), $\zeta_0 = (S_0, 0, 0, 0, 0, V_{10}, V_{20}, r_0, 0)$. The components of the DFE are given by

$$S_0 = \left(\frac{\Lambda k_6k_7k_8}{\mu k_6k_7k_8 + \mu\nu_1(\gamma_4(\mu+\nu_2+\chi)+k_8(\mu+\nu_2))}\right),$$

$$V_{10} = \frac{\nu_1S_0}{k_6}, \qquad V_{20} = \frac{\nu_1\nu_2S_0}{k_6k_7},$$

$$r_0 = \frac{\gamma_4\nu_1\nu_2S_0}{k_6k_7k_8}, \qquad N^0 = S_0 + V_{10} + V_{20} + r_0.$$

In the absence of any intervention and vaccinations, $\nu_1 = 0, \nu_2 = 0, \omega_1 = 0, \theta = 0$, implies $V_{10} = V_{20} = r_0 = 0, S_0 = \frac{\Lambda}{\mu} = N_0$. Hence, the basic reproduction number R_0 of the model (2) is given by

$$R_0 = \frac{\sigma\beta c_1}{k_2k_3k_4k_5} + \frac{\sigma\beta_1 N_0 c_2}{k_2k_3k_4k_5\epsilon}. \tag{3}$$

It is clear that

$$R_c = \frac{\sigma\beta c_1}{k_2k_3k_4k_5}\frac{S_0 + V_{10}\theta}{N^0} + \frac{\sigma\beta_1 c_2}{k_2k_3k_4k_5\epsilon}S_0.$$

Since $\frac{S_0}{N^0} \leq 1$ and $0 \leq \frac{S_0+V_{10}\theta}{N^0} \leq 1$, this implies that $R_c \leq R_0$.

Theorem 2. The DFE ζ_0 of the system (2) is locally asymptotically stable if $R_c < 1$ and unstable if $R_c > 1$.

3.3 Existence and stability of the endemic equilibria

We define the endemic equilibrium

$$\zeta^* = \left(S^*, E^*, I_a^*, I^*, I_0^*, V_1^*, V_2^*, R^*, U^*\right)$$

of system (2) as a steady state with all infected components strictly positive. From the equilibrium conditions, we obtain

$$E^* = b_1 I^*, \qquad I_a^* = b_2 I^*, \qquad I_0^* = b_3 I^*, \qquad U^* = b_4 I^*, \qquad V_2^* = \frac{\nu_2 V_1^*}{k_7},$$
$$R^* = \frac{\gamma_2 I^* + \gamma_1 I_a^* + \gamma_3 I_0^* + \gamma_4 V_2^*}{k_8},$$

where

$$N^* = S^* + E^* + I_a^* + I^* + I_0^* + V_1^* + V_2^* + R^*$$

and

$$b_1 = \frac{k_4}{k_9\sigma}, \qquad b_2 = \frac{b_1\sigma\psi}{k_3}, \qquad b_3 = \frac{b_1\sigma\phi}{k_5}, \qquad b_4 = \frac{b_2\theta_1 + b_3\theta_3 + \theta_2}{\epsilon}.$$

Substituting these expressions into the remaining equilibrium equations reduces the problem to a system of three coupled nonlinear algebraic equations in the unknowns S^*, I^* and V_1^*. The coefficients of these equations can be written in terms of auxiliary quantities such as

$$b_5 = b_2\gamma_1 + \gamma_2 + b_3\gamma_3,$$
$$b_6 = b_5k_7 + (1 + b_1 + b_2 + b_3)k_7k_8,$$
$$b_7 = k_7k_8,$$

$$
\begin{aligned}
b_8 &= k_7k_8 + k_8\nu_2 + \gamma_4\nu_2,\\
b_9 &= b_2 + \eta_1 + b_3\eta_2,\\
g_1 &\equiv \Lambda + \frac{(b_5k_7I^* + \gamma_4\nu_2V_1^*)\,\xi}{b_7} + \omega_1V_1^*\\
&\quad - \left(k_1 + b_4\beta_1I^* + \frac{b_7\beta b_9I^*}{b_6I^* + b_7S^* + b_8V_1^*}\right)S^* = 0,\\
g_2 &\equiv -b_1k_2 + b_4\beta_1S^* + \frac{b_7\beta b_9(S^* + V_1^*\theta)}{b_6I^* + b_7S^* + b_8V_1^*} = 0,\\
g_3 &\equiv \nu_1S^* - V_1^*\left(k_6 + \frac{b_7\beta b_9\theta I^*}{b_6I^* + b_7S^* + b_8V_1^*}\right) = 0.
\end{aligned}
$$

This process yields a system of three nonlinear algebraic equations in the steady-state variables S^*, I^*, and V_1^*:

$$g_1(S^*, I^*, V_1^*) = 0, \qquad g_2(S^*, I^*, V_1^*) = 0, \qquad g_3(S^*, I^*, V_1^*) = 0.$$

The resulting equations are highly nonlinear and contain rational terms, making analytical manipulation and the derivation of explicit solutions extremely challenging. Consequently, deriving closed-form expressions for the equilibrium (S^*, I^*, V_1^*) or establishing simple algebraic conditions that guarantee the existence or uniqueness of a positive endemic equilibrium is highly difficult. Therefore, the equilibrium values are computed numerically. For the parameter values given in Table 2, the control reproduction number $R_c = 1.50(> 1)$. Solving numerically, the system $(g_1, g_2, g_3) = (0, 0, 0)$ yields a unique positive endemic equilibrium $\zeta^* = \big(29503, 523, 772, 147, 23, 877, 42, 15575, 0.274\big)$.

Moreover, we study the occurrence of a transcritical bifurcation by applying the center manifold theory [8]. For convenience, we perform the change of variables

$$S = x_1 + S_0, \quad E = x_2, \quad I_a = x_3, \quad I = x_4, \quad I_0 = x_5,$$

$$V_1 = x_6 + V_{10}, \quad V_2 = x_7 + V_{20}, \quad R = x_8 + r_0, \quad U = x_9.$$

With this transformation, the model system (1) can be written in the following compact form:

$$\begin{cases} \frac{dx_1}{dt} &= \Lambda - \frac{\beta(\eta_1 x_4+x_3+\eta_2 x_5)(x_1+S_0)}{N+(S_0+V_{10}+V_{20}+r_0)} - \beta_1(x_1+S_0)x_9 + \chi(r_0+x_8) \\ & \quad +\omega_1(x_6+V_{10}) - k_1(x_1+S_0), \\ \frac{dx_2}{dt} &= \frac{\beta(\eta_1 x_4+x_3+\eta_2 x_5)((x_1+S_0)+\theta(x_6+V_{10}))}{N+(S_0+V_{10}+V_{20}+r_0)} + \beta_1(x_1+S_0)x_9 - k_2 x_2, \\ \frac{dx_3}{dt} &= \sigma\psi x_2 - k_3 x_3, \\ \frac{dx_4}{dt} &= \sigma k_9 x_2 - k_4 x_4, \\ \frac{dx_5}{dt} &= \sigma\phi x_2 - k_5 x_5, \\ \frac{dx_6}{dt} &= \nu_1(x_1+S_0) - \frac{\theta\beta(\eta_1 x_4+x_3+\eta_2 x_5)(x_6+V_{10})}{N+(S_0+V_{10}+V_{20}+r_0)} - k_6(x_6+V_{10}), \\ \frac{dx_7}{dt} &= \nu_2(x_6+V_{10}) - k_7(x_7+V_{20}), \\ \frac{dx_8}{dt} &= \gamma_2 x_4 + \gamma_1 x_3 + \gamma_3 x_5 + \gamma_4(x_7+V_{20}) - k_8(r_0+x_8), \\ \frac{dx_9}{dt} &= \theta_1 x_3 + \theta_2 x_4 + \theta_3 x_5 - \epsilon x_9. \end{cases} \tag{4}$$

At the corresponding DFE ζ_0, the Jacobian of the system (4) is expressed as follows:

$$J_{\zeta_0} = \begin{bmatrix} -k_1 & 0 & -k_{10} & -k_{10}\eta_1 & -k_{10}\eta_2 & \omega_1 & 0 & \chi & -S_0\beta_1 \\ 0 & -k_2 & k_{11} & k_{11}\eta_1 & k_{11}\eta_2 & 0 & 0 & 0 & S_0\beta_1 \\ 0 & \sigma\psi & -k_3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma k_9 & 0 & -k_4 & 0 & 0 & 0 & 0 & 0 \\ 0 & \sigma\phi & 0 & 0 & -k_5 & 0 & 0 & 0 & 0 \\ \nu_1 & 0 & -k_{12} & -k_{12}\eta_1 & -k_{12}\eta_2 & -k_6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \nu_2 & -k_7 & 0 & 0 \\ 0 & 0 & \gamma_1 & \gamma_2 & \gamma_3 & 0 & \gamma_4 & -k_8 & 0 \\ 0 & 0 & \theta_1 & \theta_2 & \theta_3 & 0 & 0 & 0 & -\epsilon \end{bmatrix},$$

where

$$k_{10} = \frac{S_0\beta_c}{(S_0+V_{10}+V_{20}+r_0)}, \quad k_{11} = \frac{\beta_c(S_0+V_{10}\theta)}{(S_0+V_{10}+V_{20}+r_0)}, \quad k_{12} = \frac{V_{10}\beta_c\theta}{(S_0+V_{10}+V_{20}+r_0)}.$$

Assuming $R_c = 1$, we determine the critical value β_c of the bifurcation parameter β, given by

$$\beta_c = -\frac{(r_0+S_0+V_{10}+V_{20})\,(-k_2k_3k_4k_5\epsilon + S_0\beta_1\sigma c_2)}{\epsilon(S_0+V_{10}\theta)\sigma c_1}.$$

It can be easily verified that the Jacobian at $\beta = \beta_c$ has a right eigenvector (corresponding to the zero eigenvalue) given by $W = (w_1, w_2, w_3, w_4, w_5, w_6, w_7, w_8, w_9)^T$, where

$$w_2 = 1, \quad w_3 = \frac{\sigma\psi}{k_3}, \quad w_4 = \frac{k_9\sigma}{k_4}, \quad w_5 = \frac{\sigma\phi}{k_5},$$

$$w_9 = \frac{\sigma c_2}{k_3k_4k_5\epsilon}, \quad w_6 = \frac{\nu_1 w_1}{k_6} - \frac{c_1k_{12}\sigma}{k_3k_4k_5k_6}, \quad w_7 = \frac{\nu_2(k_3k_4k_5w_1\nu_1 - c_1k_{12}\sigma)}{k_3k_4k_5k_6k_7},$$

$$w_1 = -\frac{\sigma(c_2k_6k_7k_8s_0\beta_1 - c_3k_6k_7\epsilon\chi + c_1\epsilon(k_{10}k_6k_7k_8 + k_{12}(\gamma_4\nu_2\chi + k_7k_8\omega_1)))}{k_3k_4k_5\epsilon(k_1k_6k_7k_8 - \nu_1(\gamma_4\nu_2\chi + k_7k_8\omega_1))},$$

$$w_8 = \frac{\sigma(-\gamma_4\nu_2(c_2s_0\beta_1\nu_1 + c_1\epsilon(k_1k_{12} + k_{10}\nu_1)) + c_3\varepsilon(k_1k_6k_7 - \nu_1(k_7\omega_1)))}{k_3k_4k_5\epsilon(k_1k_6k_7k_8 - \nu_1(\gamma_4\nu_2\chi + k_7k_8\omega_1))}.$$

The components of the left eigenvector, associated with the zero eigenvalue, are $V = (v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9)$, which must fulfill the conditions $V.J = 0$ and $V.W = 1$. Solving these, we obtain

$$v_1 = 0, \quad v_6 = 0, \quad v_7 = 0, \quad v_8 = 0, \quad v_9 = \frac{\beta_1S_0v_2}{\epsilon}, \quad v_3 = \frac{v_2(k_{11}\epsilon + \beta_1\theta_1S_0)}{k_3\epsilon},$$

$$v_4 = \frac{v_2(k_{11}\epsilon\eta_1 + \beta_1\theta_2S_0)}{k_4\epsilon}, \quad v_5 = \frac{v_2(k_{11}\epsilon\eta_2 + \beta_1\theta_3S_0)}{k_5\epsilon}, \quad v_2 = \frac{k_3^2k_4^2k_5^2\epsilon^2}{m_5},$$

where

$$\begin{aligned} m_5 &= c_2k_3k_4k_5S_0\beta_1\sigma + k_4^2k_5^2\epsilon(k_{11}\epsilon + S_0\beta_1\theta_1)\sigma\psi \\ &\quad + k_3^2\epsilon(k_5^2k_9(k_{11}\epsilon\eta_1 + S_0\beta_1\theta_2)\sigma + k_4^2(k_5^2\epsilon + k_{11}\epsilon\eta_2\sigma\phi + S_0\beta_1\theta_3\sigma\phi)), \\ m_6 &= (w_3 + w_4\eta_1 + w_5\eta_2). \end{aligned}$$

The bifurcation coefficients a and b, as given in [8, Theorem 4.1], are computed using the formulas below:

$$\begin{aligned} a &= \sum_{k,i,j=1}^{9} v_kw_iw_j\frac{\partial^2 f_k}{\partial x_i\partial x_j}(0,0) \\ &= \frac{1}{N_0^2}2(v_2)(r_0{}^2w_1w_9\beta_1 + S_0{}^2w_1w_9\beta_1 \\ &\quad + ((v_{10} + v_{20})w_1\left((v_{10} + v_{20})w_9\beta_1 + \beta_cm_6\right) \\ &\quad + r_0w_1\left(2(s_0 + v_{10} + v_{20})w_9\beta_1 + \beta_c(w_3 + w_4\eta_1 + w_5\eta_2)\right) + r_0w_6\beta_cm_6\theta \\ &\quad - (-v_{20}w_6 + v_{10}(w_1 + w_2 + w_3 + w_4 + w_5 + w_7 + w_8))\beta_cm_6\theta \\ &\quad + S_0\left(2(v_{10} + v_{20})w_1w_9\beta_1\right. \\ &\quad \left.-\beta_cm_6(w_2 + w_3 + w_4 + w_5 + w_6 + w_7 + w_8 - w_6\theta)\right). \end{aligned}$$

$$
\begin{aligned}
b &= \sum_{k,i,j=1}^{9} v_k w_i \frac{\partial^2 f_k}{\partial x_i \partial \beta}(0,0) \\
&= v_2 \left(\frac{w_3(S_0 + V_{10}\theta)}{r_0 + S_0 + V_{10} + V_{20}} + \frac{w_4\eta_1(S_0 + v_{10}\theta)}{r_0 + S_0 + V_{10} + V_{20}} + \frac{w_5\eta_2(S_0 + V_{10}\theta)}{r_0 + S_0 + V_{10} + V_{20}} \right).
\end{aligned}
$$

Given that $a < 0$ and $b > 0$ when $\beta = \beta_c$, [8, Theorem 4.1 and Remark 1] implies the presence of a transcritical bifurcation at $R_c = 1$. Consequently, the endemic equilibrium point is locally asymptotically stable for values of R_c exceeding 1.

3.4 Global stability of DFE

This section examines the global stability of the DFE.

Theorem 3. If $R_c < 1$, the DFE ζ_0 in model (2) is globally asymptotically stable, whereas it becomes unstable when $R_c > 1$.

Proof. We construct an appropriate Lyapunov function to demonstrate the global stability of the DFE ζ_0,

$$
\mathcal{G} = a_1 E + a_2 I_a + a_3 I + a_4 I_0 + a_5 U, \tag{5}
$$

The nonnegative constants a_j for $j = 1, 2, \ldots, 5$ will be specified later. Moreover, by substituting the expressions from model (2) and differentiating (5), we arrive at the following result:

$$
\begin{aligned}
\mathcal{G}' &= a_1 E' + a_2 I_a' + a_3 I' + a_4 I_0' + a_5 U' \\
&= a_1 \left(\frac{\beta(\eta_1 I + I_a + \eta_2 I_0)(S + \theta V)}{N} + \beta_1 S U - k_2 E \right) + a_2(\sigma\psi E - k_3 I_a) \\
&\quad + a_3(\sigma k_9 E - k_4 I) + a_4(\sigma\phi E - k_5 I_0) + a_5(\theta_1 I_a + \theta_2 I + \theta_3 I_0 - \epsilon U) \\
&\leq a_1 \left(\frac{\beta(\eta_1 I + I_a + \eta_2 I_0)(S_0 + \theta V_{10})}{N^0} + \beta_1 S_0 U - k_2 E \right) + a_2(\sigma\psi E - k_3 I_a) \\
&\quad + a_3(\sigma k_9 E - k_4 I) + a_4(\sigma\phi E - k_5 I_0) + a_5(\theta_1 I_a + \theta_2 I + \theta_3 I_0 - \epsilon U) \\
&= \left(-a_1 k_2 + a_2\sigma\psi + a_3 k_9 \sigma + a_4 \sigma\phi\right) E + \left(\frac{\beta a_1 (S_0 + \theta V_{10})}{N^0} - a_2 k_3 + a_5 \theta_1 \right) I_a \\
&\quad + \left(\frac{\beta a_1 \eta_1 (S_0 + \theta V_{10})}{N^0} - a_3 k_4 + a_5 \theta_2 \right) I \\
&\quad + \left(\frac{\beta a_1 \eta_2 (S_0 + \theta V_{10})}{N^0} - a_4 k_5 + a_5 \theta_3 \right) I_0 + \left(\frac{a_1(\beta_1 \Lambda)}{\mu p} - a_5 \epsilon \right) U.
\end{aligned} \tag{6}
$$

We will assign the constants the following values:

$$a_1 = \frac{k_5 N^0 \epsilon}{S_0\beta\epsilon\eta_2 + V_{10}\beta\epsilon\eta_2\theta + N^0 S_0\beta_1\theta_3}, \quad a_4 = 1, \quad a_5 = \frac{(a_1 S_0\beta_1)}{\epsilon},$$

$$N^0 = S_0 + V_{10} + V_{20} + r_0,$$

$$a_2 = \frac{(a_1 S_0\beta\epsilon + a_1 V_{10}\beta\epsilon\theta + a_1 N^0 S_0\beta_1\theta_1)}{k_3 N^0\epsilon},$$

$$a_3 = \frac{(a_1 S_0\beta\epsilon\eta_1 + a_1 V_{10}\beta\epsilon\theta\eta_1 + a_1 N^0 S_0\beta_1\theta_2)}{k_4 N^0\epsilon}.$$

Utilizing inequality (6), we obtain the result below:

$$\mathcal{G}' \leq \left(\frac{k_2 k5 N^0(R_c - 1)\epsilon}{\beta\epsilon\eta_2(S_0 + V_{10}\theta) + N^0 S_0\beta_1\theta_3}\right) E. \tag{7}$$

Hence, $\mathcal{G}' \leq 0$ whenever $R_c < 1$, and equality holds only when $E = 0$. By LaSalle's invariance principle, ζ_0 is globally asymptotically stable for $R_c < 1$. □

4 Numerical simulation

4.1 Estimation of the model parameters

In this section, we calibrate the proposed model using real-world data to gain a deeper understanding of the pandemic's transmission dynamics.

To begin, baseline parameter values are established by analyzing publicly available COVID-19 data, relevant datasets, and verified sources from the literature. Recent evidence indicates that the mean intervals between successive SARS-CoV-2 infections are approximately 204.4±51.1 days for the Alpha variant, 291.2 ± 58.2 days for the Delta variant, and 361.2 ± 131.6 days for the Omicron variant [30]. As the present study investigates the Omicron outbreak, the average period of immunity acquired following infection is taken to be $1/\chi = 361$ days. Specifically, we utilize data on the total reported COVID-19 cases in India spanning the period from 30 December 2021 to 30 April 2022 [16].

Model fitting is performed using the least squares method [25], which minimizes the discrepancies between observed data points, Y_i, and corresponding model estimates, X_i. The objective is to reduce the sum of squared errors (SSE) between the observed and predicted values, defined as

$$SSE = \sum_{i=1}^{n} (Y_i - X_i)^2.$$

In Figure 2, the continuous curve represents the fitted model, while the star-shaped markers denote the actual reported daily confirmed cases. From this calibration, the control reproduction number is estimated as $R_c = 1.17$, indicating that the Omicron variant retains a high potential for rapid transmission.

Overall, the simulated results align well with the observed data, validating the model's performance. The corresponding parameter estimates are summarized in Table 2.

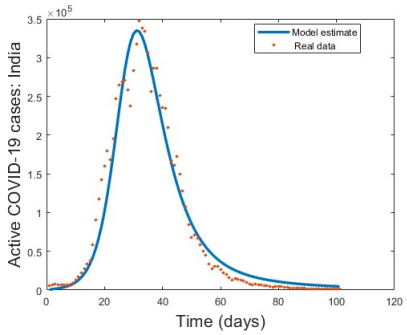


Figure 2: Fitted curve of confirmed cases in India for proposed model [16]

4.2 Sensitivity analysis

Investigating the influence of various parameters on the control reproduction number, R_c, is fundamental to understanding the behaviour of infectious diseases. This type of analysis enables the identification of key contributors to disease transmission and facilitates the formulation of effective intervention strategies during epidemics. The normalized forward sensitivity index of a parameter p with respect to R_c is expressed as (see [25])

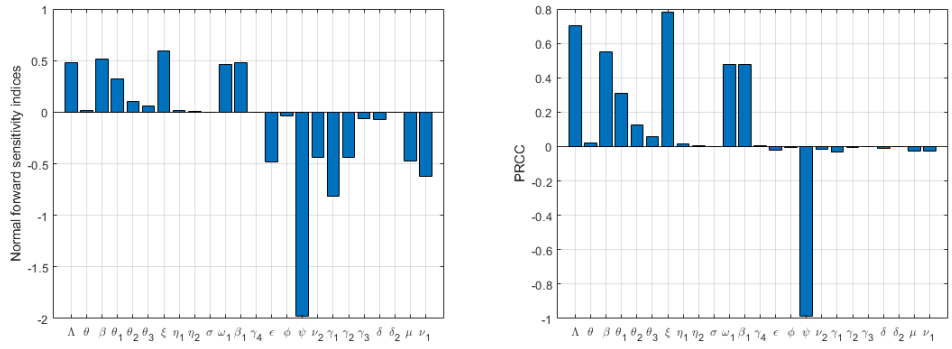
$$\Gamma_{R_c}^{p} = \frac{\partial R_c}{\partial p} \cdot \frac{p}{R_c}.$$

.

The sensitivity indices for R_c, computed using parameter values from Table 2, are listed in Table 3. The results indicate that increases in parameters such as $\beta, \omega_1, \sigma, \mu, \eta_1, \eta_2, \beta_1, \Lambda, \chi, \theta_1, \theta_2, \theta_3$, and γ_4 lead to an increase in R_c. In contrast, parameters like $\psi, \phi, \nu_1, \nu_2, \epsilon, \delta_1, \delta_2, \gamma_1, \gamma_2, \gamma_3$ are associated with a reduction in R_c. As shown in Figure 3(a), the parameter χ exerts the most

significant positive effect, whereas ψ has the most notable negative effect on R_c. Since lower values of R_c are desirable to control disease spread, it is advisable to increase the parameters with negative indices and reduce those with positive indices.

To assess how different parameters affect the infected population $I(t)$, a sensitivity analysis of the model described by (1) is performed using the PRCC method and the Latin hypercube sampling (LHS) strategy, as outlined in [**?**]. The analysis assumes a uniform distribution for all model parameters, with $I(t)$ serving as the response function. The PRCC values provide insight into the strength and direction of the relationship between each parameter and $I(t)$, thus highlighting the most influential parameters. Figure 3(b) displays the PRCC results obtained from 500 simulation runs.



(a) Normalized forward sensitivity indices $\Gamma^p_{R_c}$ (b) PRCC sensitivity indices of $I(t)$

Figure 3: Sensitivity analyses of the model parameters as in Table 1

Figure 4 presents a 2-D contour plot, while Figure 5 illustrates a 3-D contour plot of R_c as a function of vaccination rates for dose I (ν_1) and dose II (ν_2). Both figures indicate that as vaccination rates increase over time, the incidence of COVID-19 declines sharply.

4.3 Impact of vaccination

To investigate the impact of the first dose of the vaccine (Dose I) on the infected population, a numerical simulation is conducted for varying vaccination rates, denoted by ν_1, as illustrated in Figure 6(a). Similarly, the impact of the second dose on the infected population I is presented in Figure 6(b) for different levels of ν_2. Without the second dose, the number of infected individuals experiences a significant surge. However, with an increase in the administration of the second

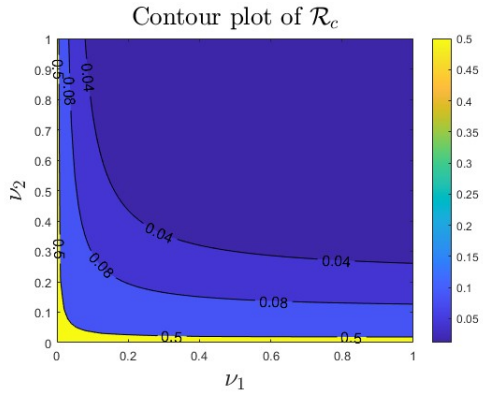


Figure 4: $2D$ contour plot of R_c as a function of ν_1 and ν_2

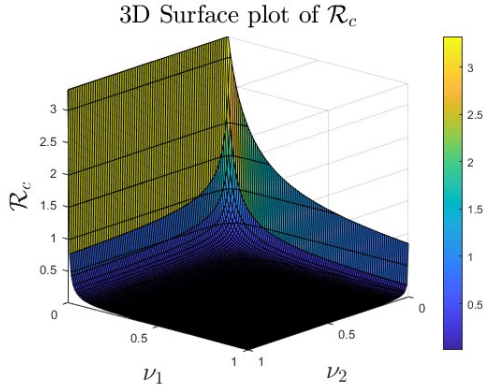
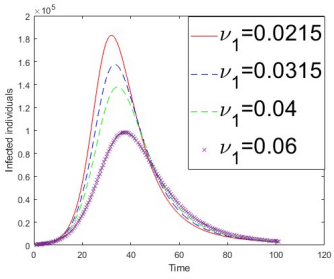


Figure 5: $3D$ contour plot of R_c as a function of ν_1 and ν_2

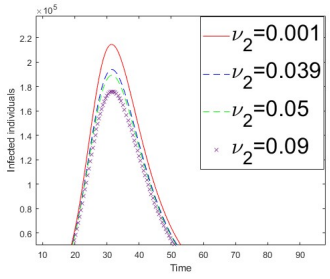
Table 3: Normalized sensitivity indices of R_c with respect to the model parameters in (1), for parameter values given in Table 2

Parameter	Sensitivity indices	Parameter	Sensitivity indices	Parameter	Sensitivity indices
Λ	0.481054	η_2	0.00559502	γ_1	-0.818185
θ	0.0185693	σ	0.0000473119	γ_4	0.00252367
β	0.518946	ω_1	0.460218	δ_1	-0.0715937
θ_1	0.318524	β_1	0.481054	ν_1	-0.624311
θ_2	0.101881	ϵ	-0.481054	γ_2	-0.0438512
θ_3	0.0606476	ϕ	-0.0369766	γ_3	-0.0656625
χ	0.588956	ψ	-1.97258	μ	-0.471738
η_1	0.0136015	ν_2	-0.436894	δ_2	-0.00056411

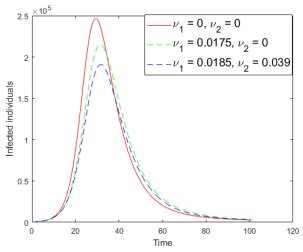
dose, there is a noticeable and sharp decline in the infected population. After receiving dose I of the vaccination, we observe a significant decrease in the infected population I, as shown in



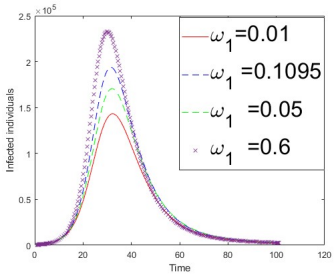
(a) Effect of ν_1 on I



(b) Effect of ν_2 on I, where $\nu_1 = 0.01$



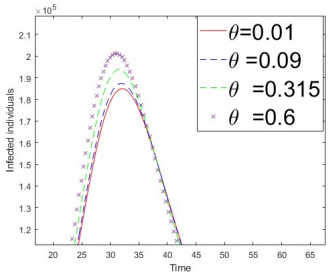
(c) Effect of ν_1 and ν_2 on I



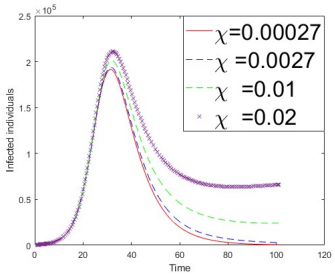
(d) Effect of ω_1 on I

Figure 6: The effects of varying ω_1, ν_1, and ν_2 on the infected population I

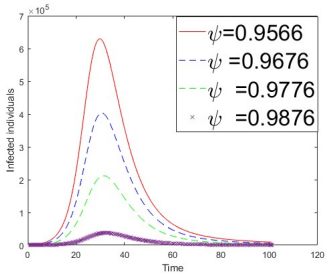
Figure 6(c). This decrease is even more pronounced after receiving dose II, compared to dose I, as shown in Figure 6(c). Asymptomatic patients have a higher rate of infectiousness, so their impact on the infection levels I is illustrated in Figure 7(c) for different levels of ψ. This is likely due to the fact that individuals in this group are often unaware of their COVID-19 infection and, as a result, may not consistently adhere to measures aimed at preventing transmission.



(a) Effect of θ on I



(b) Effect of χ on I



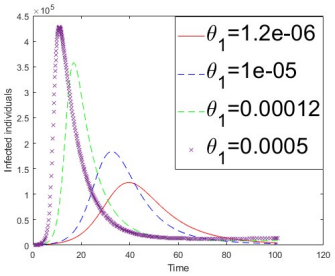
(c) Effect of ψ on I

Figure 7: The effects of varying ψ, χ, θ on the infected population I

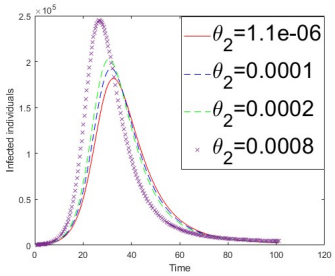
4.4 Impact of environmental contamination

Environmental contamination has been identified as a key factor contributing to the transmission of COVID-19. Studies have shown that the virus remains viable on various surfaces for differing durations—up to 4 hours on copper, 24 hours on cardboard, and as long as 72 hours on stainless steel and plastic [54]. The proposed mathematical model (1) incorporates the effects of environmental contamination, particularly from infected individuals categorized in the I_a, I, and I_0 compartments.

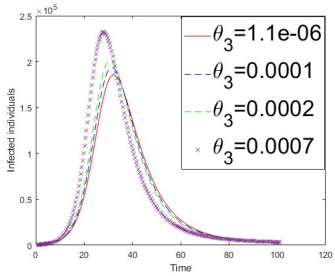
We analyzed the impact of environmental contamination on the dynamics of the proposed model. The time series in Figure 8 illustrates how parameters related to environmental contamination affect infection levels. When the rates of environmental contamination (θ_1, θ_2, and θ_3) increase, the infected population correspondingly rises, as shown in Figures 8(a)–8(c). Conversely, sanitization measures that increase ϵ reduce the number of infected individuals, as depicted in Figure 8(d). This finding suggests that eliminating the virus from contaminated environments can significantly limit the number of infections. Furthermore, a susceptible individual interacts with a contaminated environment at a transmission rate denoted by β_1. Figure 8(e) demonstrates that as β_1 increases, the infection levels rise.



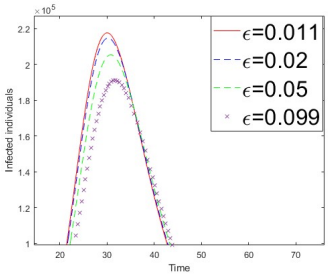
(a) Effect of θ_1 on I



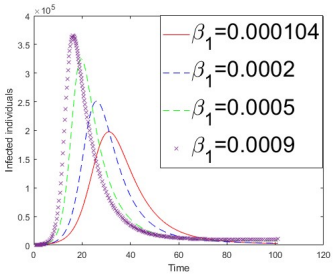
(b) Effect of θ_2 on I



(c) Effect of θ_3 on I



(d) Effect of ϵ on I



(e) Effect of β_1 on I

Figure 8: The effects of varying $\theta_1, \theta_2, \theta_3, \beta_1, \epsilon$ on the infected population I.

5 Optimal control

This section introduces an optimal control framework designed to determine the most effective control parameters. We consider the dynamical system $z'(t) = f(z(t))$ with the initial condition $z(0) = z_0$, where $z_0 \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}^n$, and $z : [0, \infty) \to \mathbb{R}^n$. To incorporate control interventions, we define control variable u_i, where $i \in \mathbb{N}$, leading to the controlled system,

$$z'(t) = f(z(t), u(t)), \quad z(0) = z_0,$$

with the control function $u : [0, \infty) \to A \subset \mathbb{R}^n$. The set of admissible controls is given by

$$X^* = \{u(t) \in L'(t_0, t_f) : u(t) \in A\}.$$

The primary objective is to determine the optimal control variables u_i that minimize the secondary transmission rate while also ensuring the associated implementation costs remain as low as possible over the time interval $t_0 \le t \le t_f$ [25]. We consider the following control variables:

1. $u_1 \equiv$ control of infections through nonpharmaceutical interventions such as face masks, hand sanitisers, etc., promoted by awareness campaigns.

2. $u_2 \equiv$ control of infections by reducing indirect transmission.

3. $u_3 \equiv$ control to enhance the uptake of the second vaccination dose.

4. $u_4 \equiv$ control to reduce virus release into the environment by encouraging infected individuals to take precautions, such as maintaining safe distances, covering coughs and sneezes with disposable tissues or clothing, and washing hands after coughing or sneezing.

5. $u_5 \equiv$ control for disinfecting contaminated environments.

To minimize both the total number of infectious cases and the corresponding implementation costs, we define the following objective function:

$$\begin{aligned} &J(u_1, u_2, u_3, u_4, u_5) \\ &= \min \int_{t_0}^{t_f} \left(Q + D_1 \frac{u_1^2}{2} + D_2 \frac{u_2^2}{2} + D_3 \frac{u_3^2}{2} + D_4 \frac{u_4^2}{2} + D_5 \frac{u_5^2}{2} \right) dt. \end{aligned} \tag{8}$$

where $Q = C_1 E + C_2 I_a + C_3 I + C_4 I_0 + C_5 U$, subject to constraints:

$$\begin{cases} \frac{dS}{dt} & = \Lambda - \frac{(1-u_1)\beta(\eta_1 I + I_a + \eta_2 I_0)S}{N} - (1-u_1-u_2)\beta_1 US - \nu_1 S + \omega_1 V_1 \\ & \quad + \chi R - \mu S, \\ \frac{dE}{dt} & = \frac{(1-u_1)\beta(\eta_1 I + I_a + \eta_2 I_0)(S+\theta V_1)}{N} + (1-u_1-u_2)\beta_1 US - \sigma E - \mu E, \\ \frac{dI_a}{dt} & = \psi\sigma E - \gamma_1 I_a - \mu I_a, \\ \frac{dI}{dt} & = (1-\psi-\phi)\,\sigma E - \gamma_2 I - \delta_1 I - \mu I, \\ \frac{dI_0}{dt} & = \phi\sigma E - \gamma_3 I_0 - \delta_2 I_0 - \mu I_0, \\ \frac{dV_1}{dt} & = \nu_1 S - (\nu_2 + \tau_1 u_3)V_1 - \omega_1 V_1 - \frac{(1-u_1)\theta\beta(\eta_1 I + I_a + \eta_2 I_0)V_1}{N} - \mu V_1, \\ \frac{dV_2}{dt} & = (\nu_2 + \tau_1 u_3)V_1 - \gamma_4 V_2 - \mu V_2, \\ \frac{dR}{dt} & = \gamma_2 I + \gamma_1 I_a + \gamma_3 I_0 + \gamma_4 V_2 - \chi R - \mu R, \\ \frac{dU}{dt} & = (1-u_4)\theta_2 I + (1-u_4)\theta_1 I_a + (1-u_4)\theta_3 I_0 - (\epsilon + \tau_2 u_5)U. \end{cases} \tag{9}$$

We begin by assuming the following initial conditions:

$$S(0) > 0, \quad E(0) \geq 0, \quad I_a(0) \geq 0, \quad I(0) \geq 0, \quad I_0(0) \geq 0, \quad V_1(0) \geq 0,$$

$$V_2(0) \geq 0, \quad R(0) \geq 0, \quad U(0) \geq 0.$$

In the optimization framework, the constants C_1, C_2, C_3, C_4, and C_5 act as weighting coefficients corresponding to the exposed individuals, the various infected groups (I_a, I, I_0), and the environmental viral concentration.

The time-dependent control parameters u_1, u_2, u_3, u_4, and u_5 are associated with the quadratic cost terms $D_1u_1^2, D_2u_2^2, D_3u_3^2, D_4u_4^2$, and $D_5u_5^2$, respectively. These squared expressions represent the severity or magnitude of each control effort.

The existence of an optimal solution to the control problem is ensured by the conditions set forth in the Filippov–Cesari theorem [2]. The Hessian matrix of the cost functional is given by

$$\mathbf{D} = \mathrm{diag}(D_1, D_2, D_3, D_4, D_5),$$

which is positive definite everywhere. Therefore, the objective functional

$$J(u_1, u_2, u_3, u_4, u_5)$$

is strictly convex. As a result, there exists a minimum value

$$D = \min\{D_1, D_2, D_3, D_4, D_5\} > 0$$

that guarantees convexity in the integrand of the objective functional

$$C_1E + C_2I_a + C_3I + C_4I_0 + C_5U + D_1\frac{u_1^2}{2} + D_2\frac{u_2^2}{2} + D_3\frac{u_3^2}{2} + D_4\frac{u_4^2}{2} + D_5\frac{u_5^2}{2}$$
$$\geq D\left(u_1^2 + u_2^2 + u_3^2 + u_4^2 + u_5^2\right),$$

This inequality is satisfied whenever $E + I_a + I + I_0 + U \geq 0$. Pontryagin's maximum principle is applied with respect to the defined state variables

$$S = S^*, \quad E = E^*, \quad I_a = I_a^*, \quad I = I^*, \quad I_0 = I_0^*, \quad V_1 = V_1^*,$$

$$V_2 = V_2^*, R = R^*, U = U^*.$$

We obtain the Hamiltonian function:

$$\begin{aligned} H =& C_1E^* + C_2I_a^* + C_3I^* + C_4I_0^* + C_5U^* + D_1\frac{u_1^2}{2} + D_2\frac{u_2^2}{2} + D_3\frac{u_3^2}{2} \\ &+ D_4\frac{u_4^2}{2} + D_5\frac{u_5^2}{2} + \lambda_1\frac{dS}{dt} + \lambda_2\frac{dE}{dt} + \lambda_3\frac{dI_a}{dt} + \lambda_4\frac{dI}{dt} + \lambda_5\frac{dI_0}{dt} \\ &+ \lambda_6\frac{dV_1}{dt} + \lambda_7\frac{dV_2}{dt} + \lambda_8\frac{dR}{dt} + \lambda_9\frac{dU}{dt}. \end{aligned}$$

The adjoint functions $\lambda_i, i = 1, 2, \ldots, 9$, obey the following equations:

$$\begin{aligned} \frac{d\lambda_1}{dt} &= -\frac{\partial H}{\partial S} \\ &= -(\lambda_6 - \lambda_2)\frac{(1-u_1)\theta\beta V_1(I_a + \eta_1 I + \eta_2 I_0)}{N^2} - (\lambda_6 - \lambda_1)\nu_1 + \mu\lambda_1 \\ &\quad - (\lambda_2 - \lambda_1)\left(U(1-u_1-u_2)\beta_1 + \frac{(1-u_1)\beta(I_a + \eta_1 I + \eta_2 I_0)}{N}\right) \\ &\quad + (\lambda_2 - \lambda_1)\left(\frac{(1-u_1)\beta S(I_a + \eta_1 I + \eta_2 I_0)}{N^2}\right), \\ \frac{d\lambda_2}{dt} &= -\frac{\partial H}{\partial E} \\ &= -(\lambda_1 - \lambda_2)\frac{(1-u_1)\beta S(I_a + \eta_1 I + \eta_2 I_0)}{N^2} \\ &\quad - (\lambda_6 - \lambda_2)\frac{(1-u_1)\theta\beta V_1(I_a + \eta_1 I + \eta_2 I_0)}{N^2} \\ &\quad - \lambda_5\sigma\phi - \lambda_3\sigma\psi - \lambda_4\sigma(1-\phi-\psi) - C_1 + \lambda_2(\mu+\sigma), \\ \frac{d\lambda_3}{dt} &= -\frac{\partial H}{\partial I_a} \\ &= -\frac{(\lambda_2 - \lambda_1)((1-u_1)\beta S(N - \eta_1 I - \eta_2 I_0))}{N^2} \end{aligned}$$

$$
\begin{aligned}
&- \frac{(\lambda_2-\lambda_6)((1-u_1)\theta\beta V_1(N-\eta_1 I-\eta_2 I_0))}{N^2}\\
&- \gamma_1\lambda_8-(1-u_4)\theta_1\lambda_9+\lambda_3(\gamma_1+\mu)-C_2,\\
\frac{d\lambda_4}{dt} &= -\frac{\partial H}{\partial I}\\
&= -\frac{(\lambda_2-\lambda_1)((1-u_1)\beta S(\eta_1 N-I_a-\eta_1 I-\eta_2 I_0))}{N^2}\\
&- \frac{(\lambda_2-\lambda_6)((1-u_1)\theta\beta V_1(\eta_1 N-I_a-\eta_1 I-\eta_2 I_0))}{N^2}\\
&- \gamma_2\lambda_8-(1-u_4)\theta_2\lambda_9+\lambda_4(\gamma_2+\delta_1+\mu)-C_3,\\
\frac{d\lambda_5}{dt} &= -\frac{\partial H}{\partial I_0}\\
&= -\frac{(\lambda_2-\lambda_1)((1-u_1)\beta S(\eta_2 N-I_a-\eta_1 I-\eta_2 I_0))}{N^2}\\
&- \frac{(\lambda_2-\lambda_6)((1-u_1)\theta\beta V_1(\eta_2 N-I_a-\eta_1 I-\eta_2 I_0))}{N^2}\\
&- \gamma_3\lambda_8-(1-u_4)\theta_3\lambda_9+\lambda_5(\gamma_3+\delta_2+\mu)-C_4,\\
\frac{d\lambda_6}{dt} &= -\frac{\partial H}{\partial V_1}\\
&= \lambda_6\mu-(\lambda_7-\lambda_6)(\nu_2+u_3\tau_1)-(\lambda_1-\lambda_6)\omega_1\\
&- (\lambda_1-\lambda_2)\frac{(1-u_1)\beta S(I_a+\eta_1 I+\eta_2 I_0)}{N^2}\\
&- (\lambda_2-\lambda_6)\frac{(1-u_1)\theta\beta(N-V_1)(I_a+\eta_1 I+\eta_2 I_0)}{N^2},\\
\frac{d\lambda_7}{dt} &= -\frac{\partial H}{\partial V_2}\\
&= -\frac{(1-u_1)\beta(I_a+\eta_1 I+\eta_2 I_0)(S(\lambda_1-\lambda_2)+V_1\theta(\lambda_6-\lambda_2))}{N^2}\\
&- \gamma_4(\lambda_8-\lambda_7)+\lambda_7\mu,\\
\frac{d\lambda_8}{dt} &= -\frac{\partial H}{\partial R}\\
&= -\frac{1}{N^2}((\lambda_1-\lambda_2)(1-u_1)\beta S(I_a+\eta_1 I+\eta_2 I_0))\\
&- \frac{1}{N^2}((\lambda_6-\lambda_2)(1-u_1)\theta\beta V_1(I_a+\eta_1 I+\eta_2 I_0))-\lambda_1\chi+\lambda_8(\mu+\chi),\\
\frac{d\lambda_9}{dt} &= -\frac{\partial H}{\partial U}\\
&= -C_5+(1-u_1-u_2)\beta_1 S(\lambda_1-\lambda_2)+\lambda_9(\epsilon+u_5\tau_2).
\end{aligned}
\tag{10}
$$

Imposing the terminal condition $\lambda_i(t_f)=0$ for $i=1,2,\ldots,9$, and using the requirement that $\frac{\partial H}{\partial u_i}=0$ for all u_i, the corresponding optimal control law is given by

$$u_1^* = \min\left\{ \max\left(0, \frac{\beta_1(\lambda_2-\lambda_1)N^*S^*U^* + \beta(I_a^* + \eta_1 I^* + \eta_2 I_0^*)(S^*(\lambda_2-\lambda_1) + \theta V_1^*(\lambda_2-\lambda_6))}{D_1 N^*}\right), u_{1\max}\right\},$$

$$\begin{aligned}
u_2^* &= \min\left\{ \max\left(0, \frac{\beta_1(-\lambda_1+\lambda_2)S^*U^*}{D_2}\right), u_{2\max}\right\},\\
u_3^* &= \min\left\{ \max\left(0, \frac{(\lambda_6-\lambda_7)\tau_1 V_1^*}{D_3}\right), u_{3\max}\right\},\\
u_4^* &= \min\left\{ \max\left(0, \frac{(\theta_1 I_a^* + \theta_2 I^* + \theta_3 I_0^*)\lambda_9}{D_4}\right), u_{4\max}\right\},\\
u_5^* &= \min\left\{ \max\left(0, \frac{\lambda_9\tau_2 U^*}{D_5}\right), u_{5\max}\right\}.
\end{aligned}$$

5.1 Numerical solution of the model under optimal control

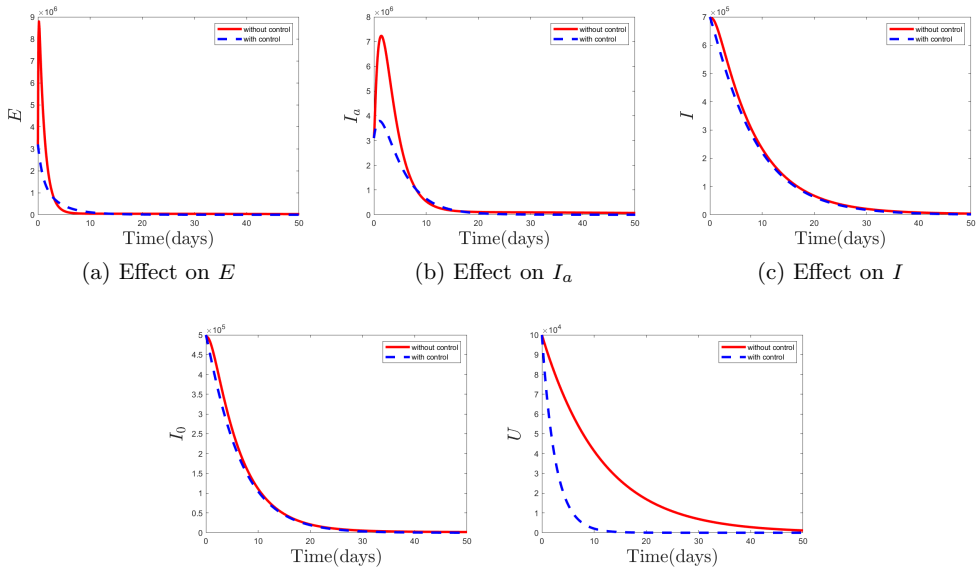
In this subsection, we conduct numerical experiments for the optimal control model to investigate how different time-dependent control strategies affect the transmission dynamics of the disease.

The simulations use the parameter set given in Table 2; several of these parameters are obtained by fitting the model to COVID-19 data reported for India. For the computations, India's total population is taken to be approximately 1.40 billion, which serves as the reference population size. To compute the optimal control, we employ the forward-backward sweep algorithm described in [2].

The trajectories displayed in Figures 9(a)–9(e) show, separately, the influence of each control variable u_i $(i = 1, 2, 3, 4, 5)$ on the state variables E, I_a, I, I_0 and U. In these plots, the solid curves represent the scenario without control, while the dotted curves indicate the system with the control strategy in place. The implementation of control measures significantly reduces both the infected population and the environmental viral load.

Figure 10 presents the optimal time profiles of the controls u_1, u_2, u_3, u_4 and u_5 and their impact on the spread of COVID-19 in the human population and the environment. In Figure 10(a), u_1 reaches its maximum value during the initial days and remains at this level until approximately $t = 90$ days. A similar pattern is observed in Figure 10(b) for u_2, which also attains an early peak and maintains this maximum level up to $t = 90$ days. In contrast, Figure 10(c) shows that u_3 reaches its highest value between 30 and 60 days and then gradually decreases until $t = 100$ days. Figure 10(d) indicates that u_4 peaks in the interval 0–50 days, followed by a

slow decline up to $t = 100$ days. Finally, Figure 10(e) reveals that u_5 also reaches its maximum between 0 and 70 days and then steadily diminishes until $t = 100$ days.



(a) Effect on E (b) Effect on I_a (c) Effect on I

(d) Effect on I_0 (e) Effect on U

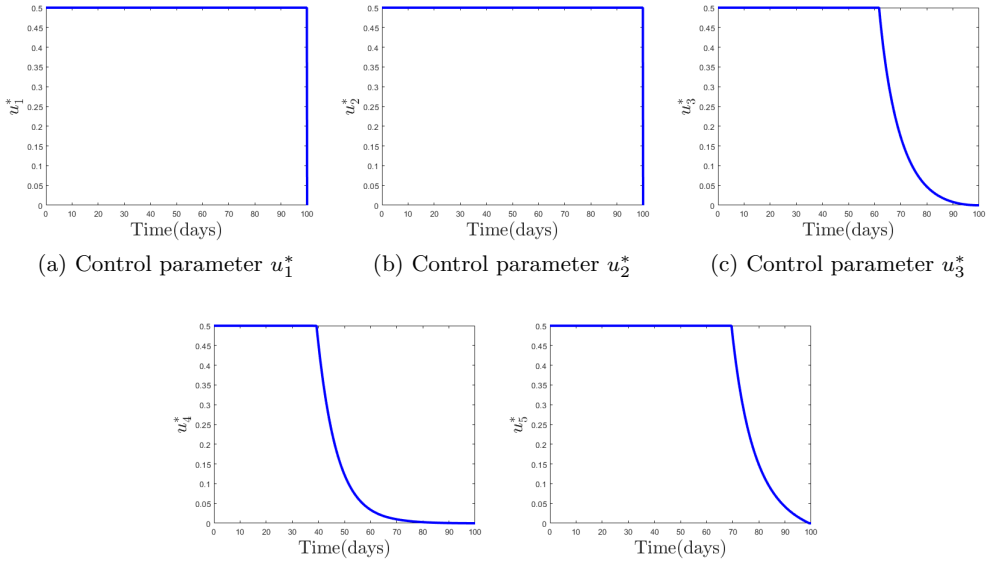
Figure 9: Effect of control measures: u_1, u_2, u_3, u_4, u_5

Table 4: Scenarios and their combination strategy

Scenario	**Strategies**
A	S-1. $(u_1 \neq 0, u_2 = 0, u_3 = 0, u_4 = 0, u_5 = 0)$
	S-2. $(u_1 = 0, u_2 \neq 0, u_3 = 0, u_4 = 0, u_5 = 0)$
	S-3. $(u_1 = 0, u_2 = 0, u_3 \neq 0, u_4 = 0, u_5 = 0)$
B	S-4. $(u_1 \neq 0, u_2 \neq 0, u_3 \neq 0, u_4 = 0, u_5 = 0)$
C	S-5. $(u_1 \neq 0, u_2 \neq 0, u_3 \neq 0, u_4 \neq 0, u_5 \neq 0)$

To achieve efficient resource allocation while reducing the economic burden associated with controlling the transmission of COVID-19, we implement selective intervention strategies based on different combinations of time-dependent control variables instead of employing all five controls simultaneously. This framework enables a comparative assessment of multiple intervention schemes (see Table 4) and facilitates the evaluation of their relative efficiency, implementation costs, and epidemiological outcomes.

Three intervention scenarios, namely Scenarios A, B, and C, are considered according to the control strategies summarized in Table 4 and illustrated in Figures 11(a)–11(e). In Scenario A,



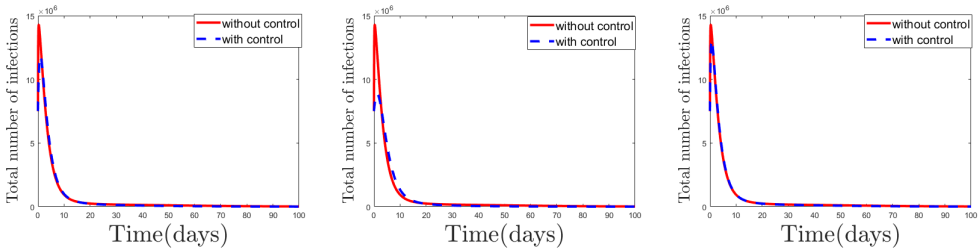
(a) Control parameter u_1^* (b) Control parameter u_2^* (c) Control parameter u_3^*

(d) Control parameter u_4^* (e) Control parameter u_5^*

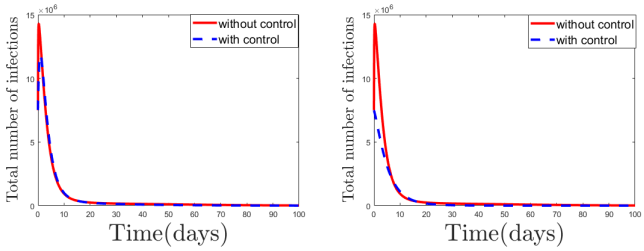
Figure 10: Control profiles: u_1, u_2, u_3, u_4 and u_5

the control measures u_1, u_2, and u_3 are evaluated individually, whereas u_4 and u_5 are excluded (Figures 11(a)–11(c)). Scenario B examines the combined implementation of strategies S-1, S-2, and S-3, with the corresponding simulation results presented in Figure 11(d). Scenario C represents the comprehensive intervention strategy in which all five control variables, u_1, u_2, u_3, u_4, and u_5, are applied simultaneously, as illustrated in Figure 11(e).

The simulation results indicate that intervention measures directed toward the infectious compartments, represented by u_1, u_2, and u_3, contribute more substantially to suppressing disease transmission than the environmental control measures u_4 and u_5. Furthermore, the simultaneous implementation of controls across all infectious classes (Figure 11(d)) produces a more pronounced reduction in disease prevalence than strategies in which these controls are applied separately (Figures 11(a)–11(c)). Among the three investigated scenarios, Scenario C consistently provides the greatest reduction in disease transmission, demonstrating that the integrated application of all available control measures is the most effective strategy for mitigating the spread of COVID-19 (Figure 11(e)).



(a) S-1: control u_1 only (b) S-2: control u_2 only (c) S-3: control u_3 only



(d) S-4: u_1, u_2 and u_3 only (e) S-5: u_1, u_2, u_3, u_4 and u_5

Figure 11: Effect of control u_1, u_2, u_3, u_4 and u_5 on total number of infections

6 Discussion and conclusion

In the present study, we developed a new mathematical model to study the transmission dynamics of COVID-19. The model focuses on the Omicron variant and incorporates individuals who have received the first and second doses of the vaccine. By integrating environmental contamination into the model, we introduce an additional layer of complexity that better represents real-world conditions, particularly in areas heavily impacted by Omicron.

The model was analyzed using dynamical systems approach, emphasizing solution positivity, boundedness, and the biological relevance of the control reproduction number derived from the next-generation matrix. Stability analysis was conducted for both the disease-free and endemic equilibria.

The numerical simulations for the proposed model, detailed in Section 4, used COVID-19 data from India, collected between December 30, 2021, and April 30, 2022, from Johns Hopkins University [16]. Figure 6(c) illustrates that vaccination plays a crucial role in shaping the dynamics of the disease. A normalized forward sensitivity analysis is conducted on the control reproduction number R_c with respect to model parameters, revealing that the transmission rate is the most significant factor. PRCC sensitivity indices are computed using the LHS technique, taking $I(t)$ as a response function with respect to model parameters. PRCC sensitivity indices

show a robust negative correlation between the infected population I and ψ (as seen in Figure 3(b)).

The analysis further demonstrated that the first dose of the vaccine provides incomplete protection, making it essential to maintain a low waning rate of immunity (ω_1) to minimize infection levels (Figure 6(d)). As ω_1 increases, the infection peak increases, indicating a greater risk of infection due to the return of vaccinated individuals to the susceptible class, which can be mitigated through the administration of a second dose. Figure 7(a) illustrates the impact of first-dose vaccine efficacy $(1-\theta)$, highlighting the need to optimize efficacy through booster doses, public health measures, and timely vaccinations to effectively manage the pandemic.

Additionally, we examined the impact of waning immunity $(1/\chi)$ on infection levels, which is critical for understanding reinfection dynamics (Figure 7(b)). Environmental contamination effects were also studied (Figure 8), with findings showing a decline in infection peaks as viral removal efforts (ϵ) increased (Figure 8(d)). This underscores the importance of safety measures, such as maintaining physical distance, covering coughs and sneezes, and practicing good hygiene to reduce community transmission.

We extended the model to incorporate an optimal control problem, examining nonpharmaceutical interventions such as face masks and hand sanitisers, reducing indirect transmission, administering the second vaccine dose, and disinfecting contaminated environments. Using Pontryagin's Maximum Principle, we established the conditions for optimal control and employed the forward-backward sweep method to assess the impact of these strategies. Numerical simulations, as shown in Figure 11(e), demonstrate that Scenario C, which combines all control variables, is the most effective in mitigating disease spread while minimizing the associated intervention costs. Overall, the results suggest that five distinct control measures should be dynamically adjusted to effectively reduce both infections and associated costs, with the second vaccine dose emerging as the most effective strategy for curbing widespread transmission.

Future research could explore extending the proposed model into a fractional-order model, as studies have shown that fractional derivatives possess additional characteristics not present in integer-order derivatives [17, 33]. Such an extension may provide a more comprehensive understanding of COVID-19 dynamics.

Funding

This study was conducted without financial support from any particular funding organization.

Conflicts of Interest

No potential conflicts of interest are associated with this research.

Data availability

All the data used in this article can be accessed freely through the website https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases#data-resources-0.

Author contributions

Amit Singh Thakur: Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Conceptualization. Govind Prasad Sahu: Writing – review & editing, Supervision. Chhatrpal Singh Sikarwar: Review & editing. Harkaran Singh: Review & editing.

References

[1] Asamoah, J.K.K., Owusu, M.A., Jin, Z., Oduro, F.T., Abidemi, A. and Gyasi, E.O. *Global stability and cost-effectiveness analysis of COVID-19 considering the impact of the environment: Using data from Ghana*, Chaos Solitons Fractals 140 (2020), 110103.

[2] Atangana, A. and Iġret, A.S. *Mathematical model of COVID-19 spread in Turkey and South Africa: Theory, methods, and applications*, Adv. Differ. Equations 2020(1) (2020), 1–89.

[3] Bagal, D.K., Rath, A., Barua, A. and Patnaik, D. *Estimating the parameters of susceptible-infected-recovered model of COVID-19 cases in India during lockdown periods*, Chaos Solitons Fractals 140 (2020), 110154.

[4] Bai, X. and Ma, S. *Stochastic dynamical behavior of COVID-19 model based on secondary vaccination*, Math. Biosci. Eng. 20(2) (2023), 2980–2997.

[5] Bajiya, V.P., Bugalia, S., Tripathi, J.P. and Martcheva, M. *Deciphering the transmission dynamics of COVID-19 in India: Optimal control and cost effective analysis*, J. Biol. Dyn. 16(1) (2022), 665–712.

[6] Bandekar, S.R., Das, T., Srivastav, A.K., Yadav, A., Kumar, A., Srivastava, P.K. and Ghosh, M. *Modeling and prediction of the third wave of COVID-19 spread in India*, Comput. Math. Biophys. 10(1) (2022), 231–248.

[7] Birkhoff, G. and Rota, G. *Ordinary differential equations*, Wiley, United Kingdom, 1978.

[8] Castillo-Chavez, C. and Song, B. *Dynamical models of tuberculosis and their applications*, Math. Biosci. Eng. 1(2) (2004), 361–404.

[9] Chavda, V.P., Bezbaruah, R., Deka, K., Nongrang, L. and Kalita, T. *The Delta and Omicron variants of SARS-CoV-2: What we know so far*, Vaccines 10(11) (2022), 1926.

[10] Davies, N.G., Abbott, S., Barnard, R.C., Jarvis, C.I., Kucharski, A.J., Munday, J.D., Pearson, C.A., Russell, T.W., Tully, D.C., Washburne, A.D. and Wenseleers, T. *Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England*, Science 372(6538) (2021), eabg3055.

[11] Fisher, A., Xu, H., He, D. and Wang, X. *Effects of vaccination on mitigating COVID-19 outbreaks: A conceptual modeling approach*, Math. Biosci. Eng. 20(3) (2023), 4816–4837.

[12] Foy, B.H., Wahl, B., Mehta, K., Shet, A., Menon, G.I. and Britto, C. *Comparing COVID-19 vaccine allocation strategies in India: A mathematical modelling study*, Int. J. Infect. Dis. 103 (2021), 431–438.

[13] Ghosh, B. *Fractional order modeling of ecological and epidemiological systems: Ambiguities and challenges*, J. Anal. 33(1) (2025), 341–366.

[14] Guo, Y. and Li, T. *Modeling the competitive transmission of the Omicron strain and Delta strain of COVID-19*, J. Math. Anal. Appl. 526(2) (2023), 127283.

[15] He, X., Chen, H., Zhu, X. and Gao, W. *Non-pharmaceutical interventions in containing COVID-19 pandemic after the roll-out of coronavirus vaccines: A systematic review*, BMC Public Health 24(1) (2024), 1–11.

[16] Humanitarian Data Exchange, *Novel coronavirus 2019 (COVID-19) cases*, https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases#data-resources-0 (accessed June 2023).

[17] Joshi, H., Jha, B.K. and Yavuz, M. *Modelling and analysis of fractional-order vaccination model for control of COVID-19 outbreak using real data*, Math. Biosci. Eng. 20(1) (2023), 213–240.

[18] Karim, S.S.A. and Karim, Q.A. *Omicron SARS-CoV-2 variant: A new chapter in the COVID-19 pandemic*, Lancet 398(10317) (2021), 2126–2128.

[19] Kar, T.K. and Ghosh, B. *Sustainability and optimal control of an exploited prey–predator system through provision of alternative food to predator*, Biosystems 109(2) (2012), 220–232.

[20] Khan, M.A. and Atangana, A. *Mathematical modeling and analysis of COVID-19: A study of new variant Omicron*, Physica A 599 (2022), 127452.

[21] Kreabkhontho, P., Teparos, W. and Theparod, T. *Potential for eliminating COVID-19 in Thailand through third-dose vaccination: A modeling approach*, Math. Biosci. Eng. 21(8) (2024), 6807–6828.

[22] Li, T. and Guo, Y. *Modeling and optimal control of mutated COVID-19 (Delta strain) with imperfect vaccination*, Chaos Solitons Fractals 156 (2022), 111825.

[23] Li, X.N., Huang, Y., Wang, W., Jing, Q.L., Zhang, C.H., Qin, P.Z., Guan, W.J., Gan, L., Li, Y.L., Liu, W.H. and Dong, H. *Effectiveness of inactivated SARS-CoV-2 vaccines against the Delta variant infection in Guangzhou: A test-negative case–control real-world study*, Emerg. Microbes Infect. 10(1) (2021), 1751–1759.

[24] Mahajan, A., Sivadas, N.A. and Solanki, R. *An epidemic model SIPHERD and its application for prediction of the spread of COVID-19 infection in India*, Chaos Solitons Fractals 140 (2020), 110156.

[25] Martcheva, M. *An introduction to mathematical epidemiology*, Springer, United States, 2015.

[26] Mohankumar, N., Rajagopal, L. and Nieto, J.J. *Optimal control for co-infection with COVID-19-associated pulmonary aspergillosis in ICU patients with environmental contamination*, Math. Biosci. Eng. 20(6) (2023), 9861–9875.

[27] Negi, S.S., Rana, P.S., Sharma, N. and Khatri, M.S. *A novel SEIAHR compartment model for accessing the impact of vaccination, intervention policies, and quarantine on the COVID-19 pandemic: A case study of most affected countries Brazil, India, Italy, and USA*, Comput. Appl. Math. 41(7) (2022), 305.

[28] Omede, B., Odionyenma, U., Ibrahim, A. and Bolaji, B. *Third wave of COVID-19: Mathematical model with optimal control strategy for reducing the disease burden in Nigeria*, Int. J. Dyn. Control 11(1) (2023), 411–427.

[29] Omorogie, E.O., Owolabi, K.M. and Olabode, B.T. *A non-linear deterministic mathematical model for investigating the population dynamics of COVID-19 in the presence of vaccination*, Healthc. Anal. 5 (2024), 1–13.

[30] Özüdoğru, O., Bahçe, Y.G. and Acer, Ö. *SARS-CoV-2 reinfection rate is higher in the Omicron variant than in the Alpha and Delta variants*, Ir. J. Med. Sci. 192(2) (2023), 751–756.

[31] Ou, J., Lan, W., Wu, X., Zhao, T., Duan, B., Yang, P., Ren, Y., Quan, L., Zhao, W., Seto, D. and Chodosh, J. *Tracking SARS-CoV-2 Omicron diverse spike gene mutations identifies multiple inter-variant recombination events*, Signal Transduct. Target. Ther. 7(1) (2022), 138.

[32] Pai, C., Bhaskar, A. and Rawoot, V. *Investigating the dynamics of COVID-19 pandemic in India under lockdown*, Chaos Solitons Fractals 138 (2020), 109988.

[33] Panwar, V.S., Uduman, P.S. and Gómez-Aguilar, J. *Mathematical modeling of coronavirus disease COVID-19 dynamics using CF and ABC non-singular fractional derivatives*, Chaos Solitons Fractals 145 (2021), 110757.

[34] Prakash, D.B., Chhetri, B., Vamsi, D.K.K., Balasubramanian, S. and Sanjeevi, C.B. *Low temperatures or high isolation delay increases the average COVID-19 infections in India: A mathematical modeling approach*, Comput. Math. Biophys. 9 (2021), 146–174.

[35] Perra, N. *Non-pharmaceutical interventions during the COVID-19 pandemic: A review*, Phys. Rep. 913 (2021), 1–52.

[36] Rana, R., Kant, R., Huirem, R.S., Bohra, D. and Ganguly, N.K. *Omicron variant: Current insights and future directions*, Microbiol. Res. 265 (2022), 127204.

[37] Sahu, G.P. and Thakur, A.S. *Mathematical modeling of COVID-19 spread with media coverage and optimal control analysis*, Iran. J. Numer. Anal. Optim. 15(3) (2025), 952–992.

[38] Sahu, G.P. and Thakur, A.S. *Analyzing the role of media, environment, and treatment limits in epidemic spread: A mathematical and optimal control approach*, Netw. Model. Anal. Health Informatics Bioinform. 14(1) (2025), 87.

[39] Sardar, T., Nadim, S.K.S., Rana, S. and Chattopadhyay, J. *Assessment of lockdown effect in some states and overall India: A predictive mathematical study on COVID-19 outbreak*, Chaos Solitons Fractals 139 (2020), 1–10.

[40] Sarkar, K., Mondal, J. and Khajanchi, S. *How do the contaminated environment influence the transmission dynamics of COVID-19 pandemic?*, Eur. Phys. J. Spec. Top. 231(18–20) (2022), 3697–3716.

[41] Senapati, A., Rana, S., Das, T. and Chattopadhyay, J. *Impact of intervention on the spread of COVID-19 in India: A model based study*, J. Theor. Biol. 523 (2021), 110711.

[42] Selvamuthu, D., Khichar, D., Kalita, P. and Jain, V. *Estimation of mortality rate of COVID-19 in India using SEIRD model*, Opsearch 60(1) (2023), 539–553.

[43] Shringi, S., Sharma, H., Rathie, P.N., Bansal, J.C. and Nagar, A. *Modified SIRD model for COVID-19 spread prediction for northern and southern states of India*, Chaos Solitons Fractals 148 (2021), 111039.

[44] Shervani, Z., Bhardwaj, D., Hasan, S., Umair, Y., Qazi, M.P., Ibbrahim, A., Vuyyuru, V.P.S.R., Siddiquie, A., Sherwani, A.A., Khan, A.A., Fatma, K., Siddiqui, S., Jamal, N. and Khan, S. *The Omicron wave in India, Mumbai, and Delhi: Prevalence and pathogenicity*, Eur. J. Med. Health Sci. 4(3) (2022), 123–130.

[45] Sheng, Y., Cui, J.-A. and Guo, S. *The modeling and analysis of the COVID-19 pandemic with vaccination and isolation: A case study of Italy*, Math. Biosci. Eng. 20(3) (2023), 5966–5992.

[46] Saha, S., Samanta, G. and Nieto, J.J. *Impact of optimal vaccination and social distancing on COVID-19 pandemic*, Math. Comput. Simul. 200 (2022), 285–314.

[47] Shan, S., Luo, S., Yang, Z., Hong, J., Su, Y., Ding, F., Fu, L., Li, C., Chen, P., Ma, J. and Shi, X. *Deep learning guided optimization of human antibody against SARS-CoV-2 variants with broad neutralization*, Proc. Nat. Acad. Sci. 119(11) (2022), e2122954119.

[48] Singh, K., Verma, A. and Lakshminarayan, M. *India's efforts to achieve 1.5 billion COVID-19 vaccinations: A narrative review*, Osong Public Health Res. Perspect. 13(5) (2022), 316.

[49] Sooknanan, J. and Comissiong, D. *Trending on social media: Integrating social media into infectious disease dynamics*, Bull. Math. Biol. 82(7) (2020), 86.

[50] Sooknanan, J. and Mays, N. *Harnessing social media in the modelling of pandemics: Challenges and opportunities*, Bull. Math. Biol. 83(5) (2021), 57.

[51] Thakur, A.S. and Sahu, G.P. *Modeling the COVID-19 dynamics with Omicron variant, non-pharmaceutical interventions, and environmental contamination*, Differ. Equations Dyn. Syst. (2025), 1–25.

[52] Ullah, S. and Khan, M.A. *Modeling the impact of non-pharmaceutical interventions on the dynamics of novel coronavirus with optimal control analysis with a case study*, Chaos Solitons Fractals 139 (2020), 110075.

[53] Van den Driessche, P. and Watmough, J. *Reproduction numbers and subthreshold endemic equilibria for compartmental models of disease transmission*, Math. Biosci. 180(1–2) (2002), 29–48.

[54] Van Doremalen, N., Bushmaker, T., Morris, D.H., Holbrook, M.G., Gamble, A., Williamson, B.N., Tamin, A., Harcourt, J.L., Thornburg, N.J., Gerber, S.I. and Lloyd-Smith, J.O. *Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1*, N. Engl. J. Med. 382(16) (2020), 1564–1567.

[55] Vivekanandhan, G., Zavareh, M.N., Natiq, H., Nazarimehr, F., Rajagopal, K. and Svetec, M. *Investigation of vaccination game approach in spreading COVID-19 epidemic model with considering the birth and death rates*, Chaos Solitons Fractals 163 (2022), 112565.

[56] Wang, X., Wang, H., Ramazi, P., Nah, K. and Lewis, M. *From policy to prediction: Forecasting COVID-19 dynamics under imperfect vaccination*, Bull. Math. Biol. 84(9) (2022), 90.

[57] World Health Organization, *Coronavirus disease (COVID-19)*, https://www.who.int/health-topics/coronavirus#tab=tab_1 (accessed July 2023).

[58] Yapışkan, D., Yurtoğlu, M., Avcı, D., İskender Eroğlu, B.B. and Bonyah, E. *A novel model for monkeypox disease: System analysis and optimal preventive strategies*, Iran. J. Sci. 47(5) (2023), 1665–1677.

Iranian Journal of Numerical Analysis and Optimization
Vol. 17, No. 1, 2027, pp 145–176
https://doi.org/10.22067/ijnao.2026.99035.1883
https://ijnao.um.ac.ir/

CC BY NC

Research Article

A modified school bullying model with a recovery compartment: Stability analysis and optimal control

A. Sasmito*, C. Miftahurrizki K H and R. Maeni

Abstract

This study proposes a compartmental dynamical model of school bullying transmission incorporating an explicit recovery mechanism through counseling and institutional intervention. The population is divided into susceptible, bully, exposed, violent, and recovered subgroups. The model's qualitative properties are established, including positivity, boundedness, equilibrium analysis, and stability conditions derived using the next-generation matrix approach. The bullying-free equilibrium is shown to be both locally and globally asymptotically stable whenever $\mathcal{R}_0 < 1$, and the endemic equilibrium is shown to exist and be locally asymptotically stable when $\mathcal{R}_0 > 1$ under explicit Routh–Hurwitz conditions. Sensitivity analysis reveals that the perpetrator-generation rate α is the primary driver of bullying persistence, whereas recovery-related interventions substantially reduce long-term prevalence. An optimal control

*Corresponding author

Received 30 May 2026; revised 1 August 2026; accepted 4 September 2026

Ayomi Sasmito
Business Mathematics, School of STEM, Prasetiya Mulya University, Tangerang, Indonesia. e-mail: ayomi.sasmito@prasetiyamulya.ac.id

Camelia Miftahurrizki K H
Business Mathematics, School of STEM, Prasetiya Mulya University, Tangerang, Indonesia. e-mail: cameliamrkh@gmail.com

Rosa Maeni
Business Mathematics, School of STEM, Prasetiya Mulya University, Tangerang, Indonesia. e-mail: maenirosa6@gmail.com

How to cite this article
Sasmito, A., Miftahurrizki K H, C. and Maeni, R., A modified school bullying model with a recovery compartment: Stability analysis and optimal control. *Iran. J. Numer. Anal. Optim.*, 2027; 17(1): 145-176. https://doi.org/10.22067/ijnao.2026.99035.1883

framework based on Pontryagin's Maximum Principle is then formulated using disciplinary and victim-protection interventions. Numerical simulations show that combining both interventions yields the largest reductions in bullying and in overall cost, indicating that integrated institutional responses outperform isolated policies.

AMS subject classifications (2020): Primary 34D20; Secondary 49J15, 92D30.

Keywords: School bullying; Compartmental model; Basic reproduction number; Sensitivity analysis; Optimal control.

1 Introduction

School bullying remains a major concern in education and public health because of its substantial impact on students' psychological well-being, academic engagement, and social development [2, 25]. Recent studies have shown that bullying involvement is strongly associated with anxiety, depression, social isolation, behavioral maladjustment, and long-term delinquent tendencies [7, 23]. Moreover, prevalence estimates vary considerably across demographic groups and school environments, suggesting that bullying is not merely an isolated behavioral incident but rather a complex social process embedded within peer interaction systems and social-ecological environments [6, 9, 21].

A growing body of evidence further indicates that bullying behavior evolves through mechanisms of peer influence and social contagion [6]. Adolescents are more likely to engage in bullying when they are surrounded by peers exhibiting similar aggressive behavior, and the probability of involvement increases according to the intensity of social interaction within friendship networks [9, 13]. From a social-ecological perspective, bullying dynamics are shaped not only by individual dispositions but also by repeated interactions, group reinforcement, classroom climate, and institutional responses [9]. These observations underscore the need for dynamic frameworks that describe how students transition between behavioral states over time.

Despite extensive empirical research on bullying prevalence, risk factors, and intervention outcomes, relatively limited attention has been given to analytical frameworks that explicitly capture long-term behavioral transitions and intervention-driven recovery processes. Most existing studies focus primarily on identifying correlates of bullying involvement. In contrast, much less attention has been devoted to understanding how bullying behavior propagates, persists, or diminishes dynamically within a school population.

Mathematical modeling provides a useful framework for investigating these issues. In particular, compartmental dynamical systems enable the analysis of threshold behavior, equilibrium stability, nonlinear interaction mechanisms, and intervention effectiveness that is difficult to observe directly from empirical studies alone [6, 18]. Such approaches have been widely used in epidemiological and social-dynamics research to examine behavioral propagation processes and optimal intervention policies [18, 16].

In the context of bullying dynamics, Crokidakis [8] proposed a compartmental model dividing the school population into susceptible, bully, exposed, and violent classes, thereby demonstrating that bullying behavior can be studied within a formal dynamical systems framework. Although this model successfully captures escalation pathways associated with bullying behavior, it does not explicitly incorporate recovery-related transitions resulting from counseling, psycho-social support, or school-based restorative intervention.

This limitation is important because recent intervention studies increasingly demonstrate that targeted school programs can reduce victimization and improve psychological outcomes among affected students [12]. Systematic reviews have shown that counseling-based and school-wide intervention strategies may substantially influence behavioral recovery and social reintegration [12, 10]. Consequently, excluding recovery dynamics may limit existing models' ability to investigate the persistence of intervention effects, relapse into susceptibility, and the long-term impact of restorative school policies on bullying prevalence.

Furthermore, although optimal control approaches have been widely applied to epidemic and behavioral-dynamics models [16], relatively few studies have examined how recovery-oriented interventions interact with the dynamics of bullying propagation and long-term control strategies. From a dynamical systems perspective, the inclusion of a recovery compartment introduces an additional recovery–susceptibility feedback mechanism that may alter the system's threshold structure and stability properties.

More recent work has extended compartmental bullying models with family education and memory effects [1], while related social-behavior models have combined next-generation-matrix stability analysis with Pontryagin-based optimal control [14, 22]. Our study differs by introducing an explicit recovery compartment with a recovery-to-susceptibility feedback and by comparing perpetrator- and victim-focused controls.

Motivated by these considerations, this study proposes a modified compartmental model of school bullying dynamics that explicitly incorporates a recovery compartment to represent counseling and school-based intervention outcomes. The model is analyzed qualitatively through positivity analysis, invariant-region analysis, equilibrium analysis, including the local and global stability of the bullying-free equilibrium and the local stability of the endemic equilibrium. In addition, an optimal control framework involving punishment-based intervention and victimization-

prevention strategies is formulated using Pontryagin's Maximum Principle to determine efficient intervention policies for reducing bullying prevalence and minimizing implementation costs.

The remainder of this paper is organized as follows. Section 2 presents the mathematical formulation and modeling assumptions. Section 3 discusses the qualitative analysis of the uncontrolled model. Section 4 develops the optimal control formulation. Section 5 presents numerical simulations and discusses intervention strategies. Finally, Section 6 concludes the paper and outlines several directions for future research.

2 Mathematical formulation and the assumptions of the model

Let $S(t)$, $B(t)$, $E(t)$, $V(t)$, and $R(t)$ denote, respectively, susceptible students, active bullying perpetrators, nonretaliating victims, retaliating victims, and recovered students at time t. The total school population is

$$N(t) = S(t) + B(t) + E(t) + V(t) + R(t).$$

The recovered class represents students who are temporarily inactive in the bullying process following counseling or intervention. This compartment is introduced to distinguish students who have recovered from those who are merely susceptible to future involvement. The proposed model extends compartmental descriptions of school bullying dynamics by incorporating recovery and possible loss of intervention effects [8, 1, 3].

The model is constructed under the following assumptions:

1. Students interact homogeneously within a single school environment. New students enter the system through the susceptible class at rate (Λ), while all students leave the system at the natural removal rate (μ).

2. Bullying-related transitions are generated by interactions between susceptible students and active perpetrators. Such interactions may lead susceptible students either to become perpetrators at rate αSB or nonretaliating victims at rate βSB.

3. Nonretaliating victims may develop aggressive responses at rate δ, while aggressively retaliating victims may de-escalate and return to the nonretaliating victim class at rate σ [8].

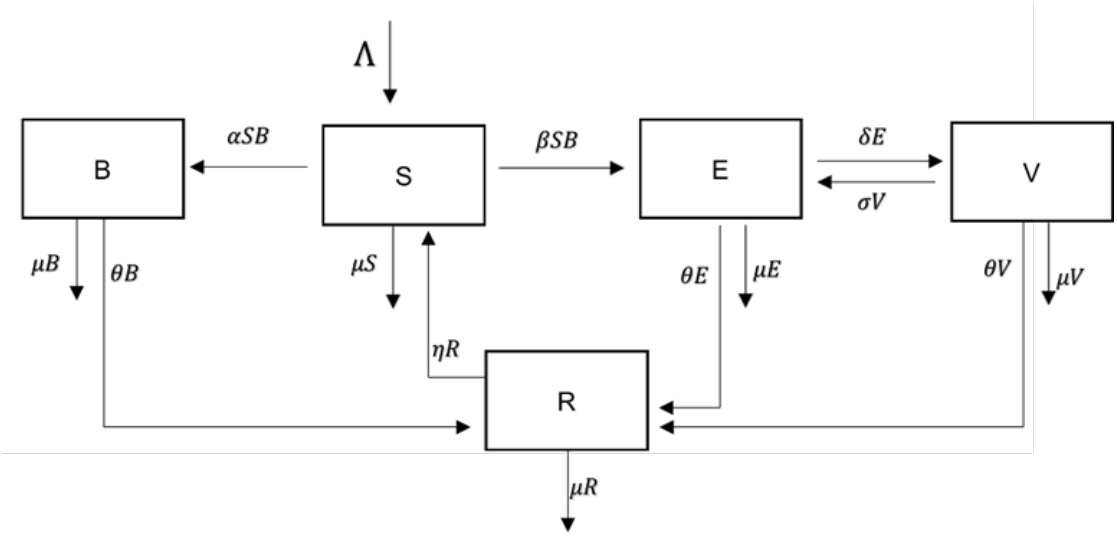


Figure 1: Transmission diagram of the bullying model with a recovery compartment.

4. Active perpetrators, nonretaliating victims, and aggressively retaliating victims may recover through counseling or school-based intervention at the common rate θ [15].

5. Recovered students are assumed to be inactive in the bullying process, although the effect of intervention may weaken over time, causing them to return to the susceptible class at rate η.

For parsimony, we assume a common recovery rate θ for perpetrators (B), nonretaliating victims (E) and retaliating victims (V). In practice these groups may respond differently to counseling and discipline, so θ should be interpreted as an average intervention effect rather than a group-specific rate. Allowing class-specific recovery rates $\theta_B, \theta_E, \theta_V$ is a straightforward extension; its effect on the threshold dynamics is discussed in Section 3.4.

The mathematical model derived from the transmission diagram in Figure 1 is given by the system of differential equations in (1). The parameters used in the proposed model are described in Table 1. Based on the assumptions stated above, we obtain the following nonlinear system of equations:

$$\begin{aligned}
\frac{dS}{dt} &= \Lambda + \eta R - \alpha SB - \beta SB - \mu S,\\
\frac{dB}{dt} &= \alpha SB - \theta B - \mu B,\\
\frac{dE}{dt} &= \beta SB + \sigma V - \delta E - \theta E - \mu E,\\
\frac{dV}{dt} &= \delta E - \sigma V - \theta V - \mu V,\\
\frac{dR}{dt} &= \theta B + \theta E + \theta V - \eta R - \mu R.
\end{aligned} \tag{1}$$

with the initial conditions given by $S(0) > 0, B(0) \geq 0, E(0) \geq 0, V(0) \geq 0$, and $R(0) \geq 0$.

Table 1: Parameters of the school bullying model (all rates are expressed per day)

Parameter	**Definition**
Λ	Recruitment rate into the school population
α	Transition rate from susceptible students to active bullying perpetrators.
β	Transition rate from susceptible students to nonretaliating victims.
δ	Progression rate from nonretaliating victims to aggressively retaliating victims.
σ	De-escalation rate from aggressive victims to nonretaliating victims.
θ	Recovery rate of individuals in B, E, and V due to counseling or intervention.
η	Return rate from recovered students to the susceptible class due to loss of intervention effects.
μ	Natural removal rate due to graduation or withdrawal.

3 Mathematical analysis of the model

Before analyzing the qualitative behavior of the system (1), we first establish that its solutions exist and are unique.

Lemma 1 (Existence and uniqueness)**.** The vector field f defined by the right-hand side of the system (1) is continuously differentiable, hence locally Lipschitz, on $\mathbb{R}^5_+$. By the Picard–Lindelöf theorem [19], for every nonnegative initial condition there exists a unique solution of (1) on a maximal forward interval. Together with the boundedness established in Lemma 2 below, this solution is defined for all $t \geq 0$.

3.1 Invariant region of the model

The following lemma establishes the boundedness of the model (1).

Lemma 2. If the initial conditions of the model (1) are within

$$\Omega = \left\{ (S, B, E, V, R) \in \mathbb{R}^5_+ : 0 \leq N(t) \leq \max\left\{ N(0), \frac{\Lambda}{\mu} \right\} \right\},$$

then all solutions of the model enter and remain in Ω.

Proof. Given the set $(S(t), B(t), E(t), V(t), R(t))$ for any solution of the system (1), and letting $N = S + B + E + V + R$, we have

$$\frac{dN}{dt} \leq \Lambda - \mu N.$$

Solving this linear differential equation yields

$$N(t) \leq \frac{\Lambda}{\mu} + \left(N_0 - \frac{\Lambda}{\mu} \right) e^{-\mu t},$$

where $N_0 = N(0)$. Since $0 < e^{-\mu t} \leq 1$ for $t \geq 0$, it follows that $0 \leq N(t) \leq \max\{N(0), \frac{\Lambda}{\mu}\}$. Therefore Ω is positively invariant.

□

3.2 Positivity of the solutions

Since the model (1) tracks the populations of students, the solutions of the model (1) have to be nonnegative. Hence, the following theorem is proved to ensure that all solutions are nonnegative.

Theorem 1. For any nonnegative initial condition in $\mathbb{R}^5_+$, the solution of system (1) remains nonnegative for all $t \geq 0$.

Proof. By Lemma 1 the right-hand side of system (1) is locally Lipschitz, so a unique solution exists on a maximal forward interval. Suppose, for contradiction, that the solution leaves the nonnegative orthant $\mathbb{R}^5_+$, and set

$$t_1 = \inf\{\, t > 0 : \min\{S(t), B(t), E(t), V(t), R(t)\} = 0 \,\}.$$

Then $t_1 > 0$ (by continuity, since all components are strictly positive at $t = 0$ except possibly B, E, V, R which are ≥ 0), and all five components are nonnegative on $[0, t_1]$ with at least one equal to zero at t_1.

For the case $S(t_1) = 0$. On $[0, t_1]$, satisfies

$$\frac{dS}{dt} = \Lambda + \eta R - (\alpha + \beta)SB - \mu S \;\geq\; -\big[(\alpha + \beta)B(t) + \mu\big]\, S,$$

because $\Lambda + \eta R \geq 0$. Integrating this differential inequality gives

$$S(t_1) \;\geq\; S(0)\, \exp\Big(-\int_0^{t_1} \big[(\alpha + \beta)B(s) + \mu\big]\, ds \Big) \;>\; 0,$$

since $S(0) > 0$ and the exponential is strictly positive. This contradicts $S(t_1) = 0$.

The same argument applies to each equation, since each has the quasi-positive form $\frac{dx_i}{dt} \geq -c_i(t)\, x_i$ on the boundary:

$$\frac{dB}{dt} \geq -(\theta + \mu)B, \quad \frac{dE}{dt} \geq -(\delta + \theta + \mu)E,$$
$$\frac{dV}{dt} \geq -(\sigma + \theta + \mu)V, \quad \frac{dR}{dt} \geq -(\eta + \mu)R,$$

where in each case the discarded terms (αSB; $\beta SB + \sigma V$; δE; $\theta B + \theta E + \theta V$) are nonnegative on $[0, t_1]$. Integrating yields $x_i(t_1) \geq x_i(0)\, e^{-\int_0^{t_1} c_i(s)\, ds} \geq 0$, and strictly positive whenever $x_i(0) > 0$, so no component can reach zero at t_1.

In every case we reach a contradiction. Hence the solution never leaves $\mathbb{R}^5_+$, i.e. $\mathbb{R}^5_+$ is positively invariant and all state variables remain nonnegative for all $t \geq 0$. □

3.3 Bullying free equilibrium point

The bullying-free equilibrium is the steady state in which bullying-related dynamics are absent in the school population. This occurs when there are no active bullies, no nonretaliating victims, and no violent responders, namely

$$B = E = V = 0.$$

In the absence of bullying involvement, no individuals enter the recovery class through intervention or counseling. Consequently, the recovered population also vanishes at equilibrium, that is, $R = 0$. The remaining population consists only of susceptible individuals. By setting the susceptible equation at steady state, one obtains

$$0 = \Lambda - \mu S,$$

which gives

$$S = \frac{\Lambda}{\mu}.$$

Thus, if the bullying-free equilibrium is denoted by $\mathcal{E}_0^* = (S_0, B_0, E_0, V_0, R_0)$, then

$$\mathcal{E}_0^* = \left(\frac{\Lambda}{\mu}, 0, 0, 0, 0\right). \tag{2}$$

3.4 Basic reproduction number

The basic reproduction number, denoted by $\mathcal{R}_0$, is defined as the expected number of perpetrators generated by a single active bully introduced into an otherwise bully-free school population during its active-bullying period [24]. It acts as a threshold: bulling cannot invade when $R_0 < 1$, whereas it can persist when $R_0 > 1$. We compute R_0 using the next-generation matrix method [24].

For the proposed model, recall the system (1) and its bullying-free equilibrium (2). To obtain $\mathcal{R}_0$, the next-generation matrix method is applied to the bullying-related compartments. Since the active bullying dynamics are represented by the compartments of bullies, nonretaliating victims, and violent victims, we take

$$x = (B, E, V)^T.$$

The terms responsible for generating new bullying-related cases are separated from the remaining transition terms. Thus, system (1), restricted to the variables (B, E, V), can be written as

$$\begin{aligned} \frac{dB}{dt} &= \alpha SB - (\theta + \mu)B, \\ \frac{dE}{dt} &= \beta SB + \sigma V - (\delta + \theta + \mu)E, \\ \frac{dV}{dt} &= \delta E - (\sigma + \theta + \mu)V. \end{aligned} \tag{3}$$

Following the next-generation matrix approach [24], we split the bullying-related subsystem into a new-appearance vector $\mathcal{F}$ and the transition vector $\mathcal{V}$, given by

$$\mathcal{F} = \begin{pmatrix} \alpha SB \\ \beta SB \\ 0 \end{pmatrix}, \qquad \mathcal{V} = \begin{pmatrix} (\theta + \mu)B \\ (\delta + \theta + \mu)E - \sigma V \\ (\sigma + \theta + \mu)V - \delta E \end{pmatrix}. \tag{4}$$

The Jacobian matrices of $\mathbb{F}$ and $\mathbb{V}$ with respect to (B, E, V), evaluated at the bullying-free equilibrium $\mathcal{E}_0^*$, are

$$\mathbb{F} = \begin{pmatrix} \frac{\alpha\Lambda}{\mu} & 0 & 0 \\ \frac{\beta\Lambda}{\mu} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \mathbb{V} = \begin{pmatrix} \theta+\mu & 0 & 0 \\ 0 & \delta+\theta+\mu & -\sigma \\ 0 & -\delta & \sigma+\theta+\mu \end{pmatrix}. \tag{5}$$

The next-generation matrix is therefore $\mathbb{F}\mathbb{V}^{-1}$. Since the first row and first column of the transition matrix are decoupled from the remaining victim-response dynamics, the dominant eigenvalue of $\mathbb{F}\mathbb{V}^{-1}$ is obtained as

$$\mathcal{R}_0 = \rho(\mathbb{F}\mathbb{V}^{-1}) = \frac{\alpha\Lambda}{\mu(\theta+\mu)}. \tag{6}$$

It is worth noting that β does not appear explicitly in $\mathcal{R}_0$, because the victimization pathway does not generate new perpetrators in the present model structure. Instead, β affects the distribution of bullying-related victims once the bully class has been established. Had the recovery rate been split into class-specific rates $\theta_B, \theta_E, \theta_V$, the basic reproduction number would read $R_0 = \alpha\Lambda/[\mu(\theta_B+\mu)]$, depending only on the perpetrator recovery rate θ_B. Hence the common-rate simplification does not affect the threshold $R_0 = 1$, and the qualitative results below remain valid; a full heterogeneous recovery analysis is left for future work.

3.5 The local stability of the bullying free-equilibrium

The following theorem establishes the local stability of the bullying-free equilibrium point.

Theorem 2. Given

$$\mathcal{R}_0 = \frac{\alpha\Lambda}{\mu(\theta+\mu)},$$

the bullying-free equilibrium point

$$\mathcal{E}_0^* = \left(\frac{\Lambda}{\mu}, 0, 0, 0, 0\right)$$

is locally asymptotically stable if $\mathcal{R}_0 < 1$ and becomes unstable if $\mathcal{R}_0 > 1$.

Proof. The local stability of the bullying-free equilibrium is determined by the eigenvalues of the Jacobian matrix of system (1) evaluated at $\mathcal{E}_0^*$.

The Jacobian matrix of system (1) is given by

$$J(S,B,E,V,R)=\begin{pmatrix} a_{11} & a_{12} & 0 & 0 & \eta \\ \alpha B & a_{22} & 0 & 0 & 0 \\ \beta B & \beta S & a_{33} & \sigma & 0 \\ 0 & 0 & \delta & a_{44} & 0 \\ 0 & \theta & \theta & \theta & a_{55} \end{pmatrix}.$$

with $a_{11}=-(\alpha+\beta)B-\mu, a_{12}=-(\alpha+\beta)S, a_{22}=\alpha S-(\theta+\mu), a_{33}=-(\delta+\theta+\mu), a_{44}=-(\sigma+\theta+\mu)$, and $a_{55}=-(\eta+\mu)$. Evaluating the Jacobian at the bullying-free equilibrium

$$\mathcal{E}_0^*=\left(\frac{\Lambda}{\mu},0,0,0,0\right),$$

we obtain

$$J(\mathcal{E}_0^*)=\begin{pmatrix} -\mu & a_{12}^0 & 0 & 0 & \eta \\ 0 & a_{22}^0 & 0 & 0 & 0 \\ 0 & a_{32}^0 & a_{33}^0 & \sigma & 0 \\ 0 & 0 & \delta & a_{44}^0 & 0 \\ 0 & \theta & \theta & \theta & a_{55}^0 \end{pmatrix}.$$

with $a_{12}^0=-\dfrac{(\alpha+\beta)\Lambda}{\mu}, a_{22}^0=\dfrac{\alpha\Lambda}{\mu}-(\theta+\mu), a_{32}^0=\dfrac{\beta\Lambda}{\mu}, a_{33}^0=-(\delta+\theta+\mu), a_{44}^0=-(\sigma+\theta+\mu)$, and $a_{55}^0=-(\eta+\mu)$. Therefore, the characteristic polynomial can be written as

$$P(\lambda)=\left[\lambda-a_{22}^0\right](\lambda-a_{55}^0)(\lambda+\mu)\left[(\lambda-a_{33}^0)(\lambda-a_{44}^0)-\sigma\delta\right].$$

Equivalently,

$$\begin{aligned} P(\lambda)=&\left[\lambda+(\theta+\mu)(1-\mathcal{R}_0)\right](\lambda+\eta+\mu)(\lambda+\mu)\\ &\times\left[\lambda^2+(\delta+\sigma+2\theta+2\mu)\lambda+(\theta+\mu)(\delta+\sigma+\theta+\mu)\right]. \end{aligned}$$

Hence, three eigenvalues are immediately identified as

$$\lambda_1=-(\eta+\mu),\qquad \lambda_2=-\mu,\qquad \lambda_3=(\theta+\mu)(\mathcal{R}_0-1).$$

Since $\eta,\mu,\theta>0$, it follows that $\lambda_1<0$ and $\lambda_2<0$. Moreover, $\lambda_3<0$ precisely when $\mathcal{R}_0<1$.

The remaining two eigenvalues are the roots of

$$\lambda^2 + (\delta + \sigma + 2\theta + 2\mu)\lambda + (\theta + \mu)(\delta + \sigma + \theta + \mu) = 0.$$

All coefficients of this quadratic polynomial are positive for nonnegative δ and σ, and positive θ and μ. Therefore, by the Routh–Hurwitz [17] criterion for a second-order polynomial, both roots have negative real parts.

Consequently, if $\mathcal{R}_0 < 1$, all eigenvalues of $J(\mathcal{E}_0^*)$ have negative real parts. By the linearization theorem [19], the bullying-free equilibrium $\mathcal{E}_0^*$ is locally asymptotically stable.

On the other hand, if $\mathcal{R}_0 > 1$, then

$$\lambda_3 = (\theta + \mu)(\mathcal{R}_0 - 1) > 0.$$

Thus, the Jacobian matrix has a positive eigenvalue, and the bullying-free equilibrium is unstable. □

Having established local stability, we now strengthen this to a global result via the Castillo-Chávez approach.

3.6 Global stability of the bullying-free equilibrium

In this section, the global asymptotic stability of the bullying-free equilibrium is investigated using the Castillo-Chávez [4, 5] approach. Let

$$X = (S, R)^T \quad \text{and} \quad Y = (B, E, V)^T,$$

where X represents the bullying-free compartments and Y represents the bullying-related classes. The model can be rewritten in the form

$$\begin{cases} \dfrac{dX}{dt} = F(X, Y), \\ \dfrac{dY}{dt} = G(X, Y), \quad G(X, 0) = 0. \end{cases} \tag{7}$$

For the proposed model, one has

$$F(X, Y) = \begin{pmatrix} \Lambda + \eta R - (\alpha + \beta)SB - \mu S \\ \theta B + \theta E + \theta V - (\eta + \mu)R \end{pmatrix},$$

and

$$G(X,Y) = \begin{pmatrix} \alpha SB - (\theta+\mu)B \\ \beta SB + \sigma V - (\delta+\theta+\mu)E \\ \delta E - (\sigma+\theta+\mu)V \end{pmatrix}.$$

Theorem 3. Given

$$\mathcal{R}_0 = \frac{\alpha\Lambda}{\mu(\theta+\mu)},$$

the bullying-free equilibrium point

$$\mathcal{E}_0^* = \left(\frac{\Lambda}{\mu}, 0, 0, 0, 0\right)$$

is globally asymptotically stable in the feasible region Ω if $\mathcal{R}_0 < 1$.

Proof. Let the feasible region be

$$\Omega = \left\{(S,B,E,V,R) \in \mathbb{R}_+^5 : S+B+E+V+R \leq \frac{\Lambda}{\mu}\right\}.$$

Since

$$\frac{d}{dt}(S+B+E+V+R) = \Lambda - \mu(S+B+E+V+R),$$

the region Ω is positively invariant.

We apply the result of Castillo-Chávez, Feng, and Huang [4], write the system as $\dot{X} = F(X,Y)$, $\dot{Y} = G(X,Y)$ with $G(X,0) = 0$, where $X = (S,R)^\top$ collects the bullying-free classes and $Y = (B,E,V)^\top$ collects the bullying-related classes. The equilibrium is globally asymptotically stable if

(H1) the reduced system $\dot{X} = F(X,0)$ has a globally asymptotically stable equilibrium $X^* = \left(\frac{\Lambda}{\mu}, 0\right)^\top$;

(H2) $\widehat{G}(X,Y) = A\,Y - G(X,Y) \geq 0$ for all $(X,Y) \in \Omega$, where $A = D_Y G(X^*, 0)$.

First, consider the reduced system

$$\frac{dX}{dt} = F(X,0),$$

which gives

$$\begin{cases} \dfrac{dS}{dt} = \Lambda + \eta R - \mu S, \\ \dfrac{dR}{dt} = -(\eta+\mu)R. \end{cases} \tag{8}$$

It is clear that $R(t) \to 0 \quad \text{as} \quad t \to \infty$. Consequently, the first equation in (8) implies

$$S(t) \to \frac{\Lambda}{\mu} \quad \text{as} \quad t \to \infty.$$

Thus, $X^* = (\Lambda/\mu, 0)^T$ is globally asymptotically stable for the reduced system.

Next, let $A = D_Y G(X^*, 0)$. Then

$$A = \begin{pmatrix} \frac{\alpha\Lambda}{\mu} - (\theta + \mu) & 0 & 0 \\ \frac{\beta\Lambda}{\mu} & -(\delta + \theta + \mu) & \sigma \\ 0 & \delta & -(\sigma + \theta + \mu) \end{pmatrix}.$$

Furthermore, $G(X, Y)$ can be written as $G(X, Y) = AY - \widehat{G}(X, Y)$, where

$$\widehat{G}(X, Y) = \begin{pmatrix} \alpha \left(\frac{\Lambda}{\mu} - S \right) B \\ \beta \left(\frac{\Lambda}{\mu} - S \right) B \\ 0 \end{pmatrix}.$$

Since $(S, B, E, V, R) \in \Omega$, we have $0 \leq S \leq \frac{\Lambda}{\mu}$. Therefore, $\widehat{G}(X, Y) \geq 0$ for all $(X, Y) \in \Omega$. It remains to observe that, if $\mathcal{R}_0 < 1$, then

$$\frac{\alpha\Lambda}{\mu} - (\theta + \mu) = (\theta + \mu)(\mathcal{R}_0 - 1) < 0.$$

The remaining two eigenvalues associated with the (E, V)-subsystem are determined by

$$\lambda^2 + (\delta + \sigma + 2\theta + 2\mu)\lambda + (\theta + \mu)(\delta + \sigma + \theta + \mu) = 0,$$

whose coefficients are positive. Hence, both roots have negative real parts.

Thus, all conditions of the Castillo-Chávez and Song theorem [5] are satisfied. Therefore, the bullying-free equilibrium

$$\mathcal{E}_0^* = \left(\frac{\Lambda}{\mu}, 0, 0, 0, 0 \right)$$

is globally asymptotically stable in Ω whenever $\mathcal{R}_0 < 1$.

□

Theorem 3 implies that, when $\mathcal{R}_0 < 1$, bullying behavior cannot persist in the school population. In this case, any small or moderate introduction of active bullies eventually dies out,

and the system converges to the bullying-free equilibrium. Biologically, this condition means that the average number of new bullies generated by one active bully during his or her active bullying period is less than unity. Therefore, the combined effects of counseling, intervention, and natural removal are sufficient to eliminate bullying dynamics from the school environment.

To simplify the notation, define the following composite parameters $a = \theta + \mu, b = \eta + \mu, c = \sigma + \theta + \mu, d = \delta + \sigma + \theta + \mu, g = \eta + \theta + \mu$, and $\Delta = \alpha\Lambda - \mu(\theta + \mu)$. These quantities are introduced only as shorthand notations for combinations of the original model parameters. The endemic equilibrium $\mathcal{E}_1^* = (S^*, B^*, E^*, V^*, R^*)$ is derived by equating the right-hand terms of the system (1) to zero, which gives

$$\mathcal{E}_1^* = \left(\frac{a}{\alpha}, \frac{b\Delta}{\mu(\alpha+\beta)g}, \frac{\beta cb\Delta}{\alpha\mu(\alpha+\beta)gd}, \frac{\beta\delta b\Delta}{\alpha\mu(\alpha+\beta)gd}, \frac{\theta\Delta}{\alpha\mu g} \right).$$

Since all parameters are assumed to be positive, the endemic equilibrium is feasible if and only if $\Delta > 0$ or equivalently $\alpha\Lambda - \mu(\theta + \mu) > 0$. Using the basic reproduction number $\mathcal{R}_0 = \frac{\alpha\Lambda}{\mu(\theta+\mu)}$, the above condition can be written as $\mathcal{R}_0 > 1$. Therefore, the endemic equilibrium exists in the positive feasible region whenever $\mathcal{R}_0 > 1$.

Proposition 1. System (1) admits a unique endemic equilibrium

$$\mathcal{E}_1^* = \left(\frac{a}{\alpha}, \frac{b\Delta}{\mu(\alpha+\beta)g}, \frac{\beta cb\Delta}{\alpha\mu(\alpha+\beta)gd}, \frac{\beta\delta b\Delta}{\alpha\mu(\alpha+\beta)gd}, \frac{\theta\Delta}{\alpha\mu g} \right)$$

with all components strictly positive if and only if $\mathcal{R}_0 > 1$, or equivalently $\Delta > 0$.

3.7 Local stability of the endemic equilibrium

The existence of the endemic equilibrium indicates the persistence of bullying in the population. Therefore, it is important to examine whether solutions of the system remain close to this equilibrium under small perturbations. This can be achieved by studying the eigenvalues of the Jacobian matrix evaluated at $\mathcal{E}_1^*$. The local stability condition is then derived from the characteristic polynomial associated with $J(\mathcal{E}_1^*)$. The result is summarized in the following theorem.

Theorem 4. Assume that all parameters are positive and that the endemic equilibrium

$$\mathcal{E}_1^* = \left(\frac{a}{\alpha}, \frac{b\Delta}{\mu(\alpha+\beta)g}, \frac{\beta cb\Delta}{\alpha\mu(\alpha+\beta)gd}, \frac{\beta\delta b\Delta}{\alpha\mu(\alpha+\beta)gd}, \frac{\theta\Delta}{\alpha\mu g} \right)$$

exists. Then the endemic equilibrium $\mathcal{E}_1^*$ is locally asymptotically stable if the following Routh–Hurwitz [17] conditions hold: $\psi_1 > 0, \psi_2 > 0, \psi_3 > 0$, and $\psi_1\psi_2 > \psi_3$, where

$$\psi_1 = p_1 + \eta + \mu, \quad \psi_2 = p_0 + (\eta + \mu)p_1, \quad \text{and} \quad \psi_3 = (\eta + \mu)p_0 - \eta\theta\frac{b\Delta}{\mu g},$$

with

$$p_1 = \frac{b\Delta}{\mu g} + 2\mu + \theta - a, \quad \text{and} \quad p_0 = \left(\frac{b\Delta}{\mu g} + \mu\right)(\theta + \mu) - a\mu.$$

Proof. The Jacobian matrix evaluated at the endemic equilibrium $\mathcal{E}_1^*$ is given by

$$J(\mathcal{E}_1^*) = \begin{pmatrix} a_{11}^1 & a_{12}^1 & 0 & 0 & \eta \\ \alpha B^* & a_{22}^1 & 0 & 0 & 0 \\ \beta B^* & \beta S^* & a_{33}^1 & \sigma & 0 \\ 0 & 0 & \delta & a_{44}^1 & 0 \\ 0 & \theta & \theta & \theta & a_{55}^1 \end{pmatrix},$$

where

$$\begin{aligned} a_{11}^1 &= -(\alpha + \beta)B^* - \mu, \quad a_{12}^1 = -(\alpha + \beta)S^*, \\ a_{22}^1 &= \alpha S^* - (\theta + \mu), \quad a_{33}^1 = -(\delta + \theta + \mu), \\ a_{44}^1 &= -(\sigma + \theta + \mu), \quad a_{55}^1 = -(\eta + \mu). \end{aligned}$$

The characteristic equation is

$$\det\left(\lambda I - J(\mathcal{E}_1^*)\right) = 0.$$

A direct computation gives

$$\det\left(\lambda I - J(\mathcal{E}_1^*)\right) = (\lambda + \theta + \mu)(\lambda + \delta + \sigma + \theta + \mu)P_3(\lambda),$$

where

$$P_3(\lambda) = (\lambda + \eta + \mu)\left(\lambda^2 + p_1\lambda + p_0\right) - \eta\theta\frac{b\Delta}{\mu g}.$$

Here

$$p_1 = \frac{b\Delta}{\mu g} + 2\mu + \theta - a, \quad \text{and} \quad p_0 = \left(\frac{b\Delta}{\mu g} + \mu\right)(\theta + \mu) - a\mu.$$

Hence, two eigenvalues are $\lambda_1 = -(\theta + \mu), \lambda_2 = -(\delta + \sigma + \theta + \mu)$. Since all parameters are positive, both eigenvalues are negative. It remains to analyze the cubic polynomial $P_3(\lambda)$. Expanding $P_3(\lambda)$ gives

$$P_3(\lambda) = \lambda^3 + \psi_1\lambda^2 + \psi_2\lambda + \psi_3,$$

where

$$\psi_1 = p_1 + \eta + \mu, \quad \psi_2 = p_0 + (\eta + \mu)p_1, \quad \psi_3 = (\eta + \mu)p_0 - \eta\theta\frac{b\Delta}{\mu g}.$$

By the Routh–Hurwitz criterion [17], all roots of the cubic polynomial $P_3(\lambda)$ have negative real parts if and only if $\psi_1 > 0, \psi_2 > 0, \psi_3 > 0$, and $\psi_1\psi_2 > \psi_3$. Therefore, under these conditions, all eigenvalues of $J(\mathcal{E}_1^*)$ have negative real parts. Hence, the endemic equilibrium $\mathcal{E}_1^*$ is locally asymptotically stable.

□

3.8 Sensitivity analysis

The normalized forward sensitivity index is used to measure the relative change in the basic reproduction number $\mathcal{R}_0$ with respect to changes in each model parameter. The normalized forward sensitivity index of $\mathcal{R}_0$ with respect to a parameter p is defined as

$$\Upsilon_p^{\mathcal{R}_0} = \frac{\partial \mathcal{R}_0}{\partial p} \times \frac{p}{\mathcal{R}_0}.$$

Here, p denotes any parameter involved in the expression of $\mathcal{R}_0$. Since the basic reproduction number is given by

$$\mathcal{R}_0 = \frac{\alpha\Lambda}{\mu(\theta + \mu)},$$

the parameters that directly influence $\mathcal{R}_0$ are α, Λ, μ, and θ.

$$\begin{aligned}
\Upsilon_\alpha^{\mathcal{R}_0} &= \frac{\partial \mathcal{R}_0}{\partial \alpha} \times \frac{\alpha}{\mathcal{R}_0} = 1 \\
\Upsilon_\Lambda^{\mathcal{R}_0} &= \frac{\partial \mathcal{R}_0}{\partial \Lambda} \times \frac{\Lambda}{\mathcal{R}_0} = 1 \\
\Upsilon_\theta^{\mathcal{R}_0} &= \frac{\partial \mathcal{R}_0}{\partial \theta} \times \frac{\theta}{\mathcal{R}_0} = -\frac{\theta}{\theta + \mu} < 0 \\
\Upsilon_\mu^{\mathcal{R}_0} &= \frac{\partial \mathcal{R}_0}{\partial \mu} \times \frac{\mu}{\mathcal{R}_0} = -\frac{\theta + 2\mu}{\theta + \mu} < 0.
\end{aligned}$$

Therefore, the sensitivity indices of $\mathcal{R}_0$ are summarized in Table 2.

The results show that the parameters α and Λ have positive sensitivity indices. This indicates that an increase in either α or Λ will increase the value of $\mathcal{R}_0$. In particular, a 1% increase in α or Λ results in a 1% increase in $\mathcal{R}_0$, provided that the other parameters remain constant.

Table 2: Sensitivity indices of $\mathcal{R}_0$ with respect to the model parameters

Parameter	**Sensitivity index**
α	1
Λ	1
θ	$-\dfrac{\theta}{\theta+\mu} < 0$
μ	$-\dfrac{\theta+2\mu}{\theta+\mu} < 0$

On the other hand, the parameters θ and μ have negative sensitivity indices. This implies that increasing either θ or μ will reduce the value of $\mathcal{R}_0$. Hence, θ and μ play an important role in reducing bullying transmission, since larger values of these parameters contribute to a lower basic reproduction number.

To further illustrate the influence of each parameter on the basic reproduction number, we present the variation of $\mathcal{R}_0$ with respect to the parameters Λ, α, θ, and μ. The threshold value $\mathcal{R}_0 = 1$ separates the nonendemic region from the endemic region. In particular, the bullying dynamics tend to disappear when $\mathcal{R}_0 < 1$, whereas they persist in the population when $\mathcal{R}_0 > 1$.

Figure 2 shows that $\mathcal{R}_0$ increases as Λ and α increase. This agrees with the positive sensitivity indices of these two parameters. Biologically, a larger recruitment rate or transmission rate increases the potential for bullying spread in the population.

In contrast, $\mathcal{R}_0$ decreases as θ and μ increase. This is consistent with their negative sensitivity indices. Hence, increasing the values of θ and μ can contribute to reducing the basic reproduction number below the threshold value $\mathcal{R}_0 = 1$. The critical values shown in the figure indicate the parameter values at which the system changes from the endemic region to the nonendemic region, or conversely.

Notably, the parameters β, δ, σ, and η do not appear in $\mathcal{R}_0$. This structural property implies that interventions directed exclusively at the victimization pathway (without addressing perpetrator formation) are insufficient to drive the system below the persistence threshold. Because $\mathcal{R}_0$ is insensitive to the victimization pathway, we next ask how intervention effort should be allocated over time, which motivates the optimal-control formulation in Section 4.

4 An optimal control model

In this section, the mathematical model of bullying in (1) is further extended by incorporating optimal control theory as a strategic approach to reduce the spread of bullying behavior in

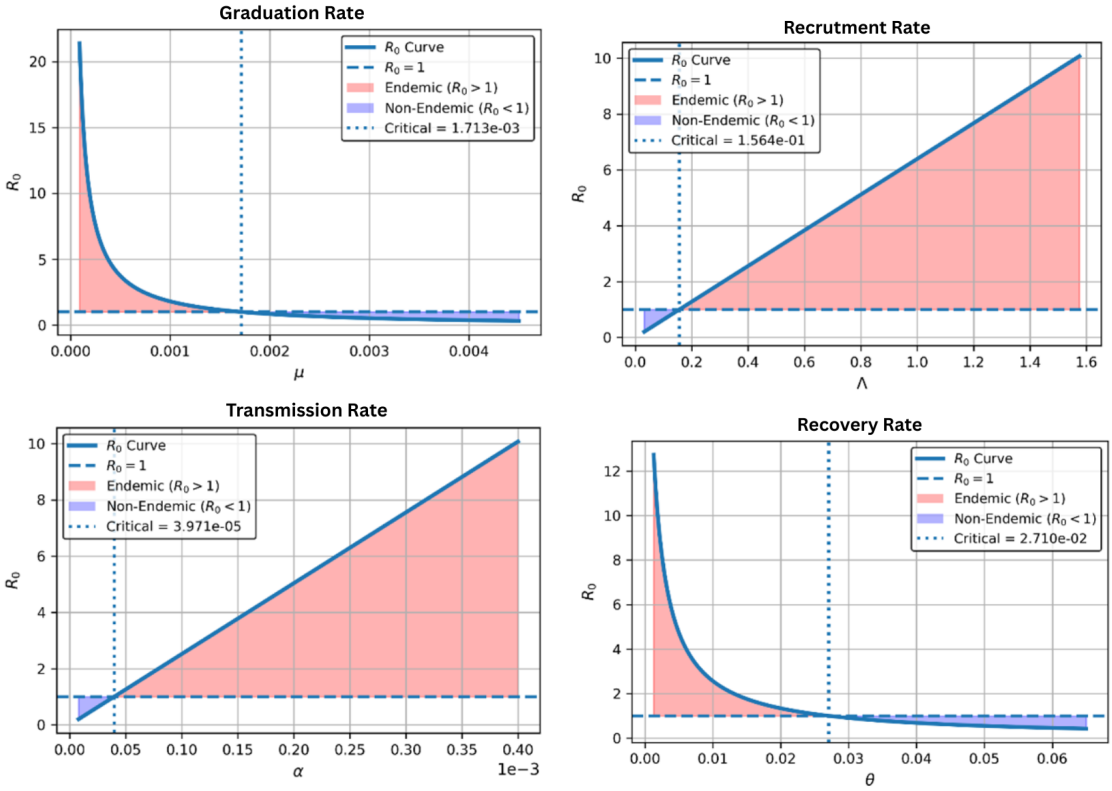


Figure 2: Graphical sensitivity analysis of $\mathcal{R}_0$ with respect to key model parameters.

the school environment. The application of control theory is intended to determine effective intervention strategies that minimize the number of perpetrators and victims while increasing the number of students who remain safe or recover from the effects of bullying.

From the perspective of applied mathematics, the control problem is formulated by introducing two time-dependent control variables, namely $u_1(t)$ and $u_2(t)$. The control variable $u_1(t)$ represents a punishment-based intervention, such as sanctions or disciplinary actions given to bullying perpetrators. Meanwhile, the control variable $u_2(t)$ represents a victimization-prevention intervention, such as teacher supervision, peer support, and anti-bullying programs designed to reduce the risk of susceptible students becoming nonretaliating victims.

Therefore, the controlled mathematical model with the implementation of $u_1(t)$ and $u_2(t)$ is given as follows:

$$\begin{aligned}
\frac{dS}{dt} &= \Lambda + \eta R - (1-u_1)\alpha SB - (1-u_2)\beta SB - \mu S,\\
\frac{dB}{dt} &= (1-u_1)\alpha SB - \theta B - \mu B,\\
\frac{dE}{dt} &= (1-u_2)\beta SB + \sigma V - \delta E - \theta E - \mu E,\\
\frac{dV}{dt} &= \delta E - \sigma V - \theta V - \mu V,\\
\frac{dR}{dt} &= \theta B + \theta E + \theta V - \eta R - \mu R.
\end{aligned} \tag{9}$$

The objective of the optimal control problem is to minimize the numbers of bullying perpetrators and students at risk while also minimizing the implementation costs of the control strategies. Therefore, the objective functional is defined as

$$J(u_1, u_2) = \int_0^{t_f} \left[A_1 B(t) + A_2 E(t) + \frac{1}{2} C_1 u_1^2(t) + \frac{1}{2} C_2 u_2^2(t) \right] dt, \tag{10}$$

where t_f is the final time, A_1 and A_2 are positive weight constants associated with bullying perpetrators and students at risk, respectively, while C_1 and C_2 are positive weight constants representing the costs of implementing the control strategies $u_1(t)$ and $u_2(t)$. Next, we investigate the existence of an optimal control for the above-mentioned problem using the result of Fleming and Rishel [11].

Theorem 5 (Existence of an optimal control)**.** There exists an optimal control pair $u^* = (u_1^*, u_2^*) \in \mathcal{U}$ such that

$$J(u_1^*, u_2^*) = \min_{(u_1,u_2)\in\mathcal{U}} J(u_1, u_2), \tag{11}$$

where the admissible control set $\mathcal{U}$ is defined by

$$\mathcal{U} = \{(u_1, u_2) \mid u_i(t) \text{ is measurable on } [0, t_f], \quad 0 \le u_i(t) \le 1, \quad i = 1, 2\}.$$

The optimal control problem is subject to the controlled system (9) with nonnegative initial conditions.

Proof. To prove the existence of an optimal control, we follow the standard existence result in optimal control theory [11]. It is sufficient to verify the following conditions:

1. The control set $\mathcal{U}$ is nonempty, closed, and convex.

2. The state variables are nonnegative and bounded.

3. The right-hand side of the state system is bounded by a linear function of the state and control variables.

4. The integrand of the objective functional is convex with respect to the control variables on $\mathcal{U}$.

5. There exist constants $d_1 > 0$, $d_2 > 0$, and $c > 1$ such that the integrand satisfies a lower bound of the form
$$L(B, E, u_1, u_2) \geq d_1 \left(|u_1|^2 + |u_2|^2\right)^{c/2} - d_2.$$
where $L(B, E, u_1, u_2) = A_1 B + A_2 E + \frac{1}{2} C_1 u_1^2 + \frac{1}{2} C_2 u_2^2$ is the integrand of the objective functional (10).

First, the control set $\mathcal{U}$ is nonempty since the pair $(u_1, u_2) = (0, 0)$ belongs to $\mathcal{U}$. Moreover, since $0 \leq u_i(t) \leq 1$ for $i = 1, 2$, the set $\mathcal{U}$ is closed, bounded, and convex.

Next, for nonnegative initial conditions, the state variables remain nonnegative for all $t \in [0, t_f]$. Indeed, on the boundary of the nonnegative orthant, the right-hand side of the controlled system satisfies
$$\left.\frac{dS}{dt}\right|_{S=0} = \Lambda + \eta R \geq 0, \quad \left.\frac{dB}{dt}\right|_{B=0} = 0,$$
$$\left.\frac{dE}{dt}\right|_{E=0} = (1 - u_2)\beta S B + \sigma V \geq 0, \quad \left.\frac{dV}{dt}\right|_{V=0} = \delta E \geq 0,$$
and
$$\left.\frac{dR}{dt}\right|_{R=0} = \theta B + \theta E + \theta V \geq 0.$$
Thus, the nonnegative orthant is positively invariant.

Let
$$N(t) = S(t) + B(t) + E(t) + V(t) + R(t).$$
Adding all equations in the controlled system (9) gives
$$\frac{dN}{dt} = \Lambda - \mu N.$$
Hence,
$$N(t) \leq \max\left\{N(0), \frac{\Lambda}{\mu}\right\},$$
which implies that all state variables are bounded on $[0, t_f]$.

Furthermore, the right-hand side of the controlled system (9) is continuously differentiable with respect to the state variables and controls. Since the state variables are bounded and $0 \leq u_i(t) \leq 1$, the right-hand side of the system is bounded by a linear function of the state and control variables.

Recall the objective functional (10), whose integrand is

$$L(B, E, u_1, u_2) = A_1 B + A_2 E + \frac{1}{2} C_1 u_1^2 + \frac{1}{2} C_2 u_2^2.$$

Since $C_1 > 0$ and $C_2 > 0$, the function L is convex with respect to u_1 and u_2. Moreover, because $B(t) \geq 0$ and $E(t) \geq 0$, we have

$$L(B, E, u_1, u_2) \geq \frac{1}{2} C_1 u_1^2 + \frac{1}{2} C_2 u_2^2.$$

Therefore,

$$L(B, E, u_1, u_2) \geq d_1 \left(|u_1|^2 + |u_2|^2 \right) - d_2,$$

where

$$d_1 = \frac{1}{2} \min\{C_1, C_2\} \qquad \text{and} \qquad d_2 = 0.$$

Thus, all the required conditions for the existence of an optimal control are satisfied. Therefore, there exists an optimal control pair $(u_1^*, u_2^*) \in \mathcal{U}$ such that

$$J(u_1^*, u_2^*) = \min_{(u_1, u_2) \in \mathcal{U}} J(u_1, u_2).$$

This completes the proof. □

To characterise the optimal controls, we apply Pontryagin's maximum principle [20]. The Hamiltonian ($\mathcal{H}$) of the optimal control problem (9)–(10) is as follows:

$$\begin{aligned} H = {} & A_1 B + A_2 E + \frac{1}{2} C_1 u_1^2 + \frac{1}{2} C_2 u_2^2 \\ & + \lambda_1 \left[\Lambda + \eta R - (1 - u_1) \alpha S B - (1 - u_2) \beta S B - \mu S \right] \\ & + \lambda_2 \left[(1 - u_1) \alpha S B - \theta B - \mu B \right] \\ & + \lambda_3 \left[(1 - u_2) \beta S B + \sigma V - \delta E - \theta E - \mu E \right] \\ & + \lambda_4 \left[\delta E - \sigma V - \theta V - \mu V \right] \\ & + \lambda_5 \left[\theta B + \theta E + \theta V - \eta R - \mu R \right]. \end{aligned}$$

where λ_i, $i = 1, 2, \ldots, 5$, are the adjoint variable functions to be determined.

Applying Pontryagin's maximum principle [20] to the Hamiltonian H yields the following adjoint system and optimality conditions. The adjoint variables $\lambda_i(t)$, $i = 1, \ldots, 5$, satisfy $\frac{d\lambda_i}{dt} = -\frac{\partial H}{\partial x_i}$, $x_i \in \{S, B, E, V, R\}$, that is,

$$\begin{aligned}
\frac{d\lambda_1}{dt} &= -\frac{\partial H}{\partial S} = \lambda_1 \left[(1-u_1)\alpha B + (1-u_2)\beta B + \mu\right] \\
&\quad - \lambda_2 \left(1-u_1\right)\alpha B - \lambda_3(1-u_2)\beta B, \\
\frac{d\lambda_2}{dt} &= -\frac{\partial H}{\partial B} = -A_1 + \lambda_1 \left[(1-u_1)\alpha S + (1-u_2)\beta S\right] \\
&\quad - \lambda_2 \left[(1-u_1)\alpha S - \theta - \mu\right] - \lambda_3(1-u_2)\beta S \\
&\quad - \lambda_5\theta, \\
\frac{d\lambda_3}{dt} &= -\frac{\partial H}{\partial E} = -A_2 + \lambda_3 \left[\delta + \theta + \mu\right] - \lambda_4\delta - \lambda_5\theta, \\
\frac{d\lambda_4}{dt} &= -\frac{\partial H}{\partial V} = -\lambda_3\sigma + \lambda_4 \left[\sigma + \theta + \mu\right] - \lambda_5\theta, \\
\frac{d\lambda_5}{dt} &= -\frac{\partial H}{\partial R} = -\lambda_1\eta + \lambda_5 \left[\eta + \mu\right],
\end{aligned}$$

together with the transversality conditions $\lambda_i(t_f) = 0, i = 1, 2, \ldots, 5$. The optimality conditions follow from $\frac{\partial H}{\partial u_1} = 0$ and $\frac{\partial H}{\partial u_2} = 0$. Since

$$\frac{\partial H}{\partial u_1} = C_1 u_1 + \alpha SB(\lambda_1 - \lambda_2), \qquad \frac{\partial H}{\partial u_2} = C_2 u_2 + \beta SB(\lambda_1 - \lambda_3),$$

we obtain

$$u_1 = \frac{\alpha SB(\lambda_2 - \lambda_1)}{C_1}, \qquad u_2 = \frac{\beta SB(\lambda_3 - \lambda_1)}{C_2}.$$

Because the controls are bounded, $0 \le u_i(t) \le 1$, projecting these expressions onto the admissible set U gives the characterisation of the optimal controls

$$\begin{aligned}
u_1^*(t) &= \min\Big\{1,\ \max\Big\{0,\ \tfrac{\alpha S^* B^* (\lambda_2 - \lambda_1)}{C_1}\Big\}\Big\}, \\
u_2^*(t) &= \min\Big\{1,\ \max\Big\{0,\ \tfrac{\beta S^* B^* (\lambda_3 - \lambda_1)}{C_2}\Big\}\Big\}.
\end{aligned}$$

The optimality system is completed by the initial conditions for the state variables and the transversality conditions for the adjoint variables. The initial conditions are given by $S(0) = S_0, B(0) = B_0, E(0) = E_0, V(0) = V_0, R(0) = R_0$. Since the terminal states are free and there is no terminal cost in the objective functional, the transversality conditions are

$$\lambda_i(t_f) = 0, \qquad i = 1, 2, \ldots, 5.$$

5 Numerical simulations and discussion

The optimality system was solved numerically using the *forward–backward sweep method* (FBSM) [16]. The state system was integrated forward in time using a fourth-order Runge–Kutta (RK4) scheme, while the adjoint system was integrated backward from the transversality conditions. All numerical computations and figures were produced in Python.

At each iteration, the controls were updated via the convex combination

$$u_i^{(k+1)} = \rho\,\hat{u}_i^{(k)} + (1-\rho)\,u_i^{(k)}, \qquad i = 1, 2,$$

where k indicates the iteration number, $\hat{u}_i^{(k)}$ denotes the control computed from the optimality conditions, and $\rho = 0.5$ is the relaxation parameter. Convergence was declared when

$$\max\left\{\left\|u_1^{(k+1)} - u_1^{(k)}\right\|_2,\ \left\|u_2^{(k+1)} - u_2^{(k)}\right\|_2\right\} < 10^{-6},$$

with a maximum of 200 iterations.

The terminal time was set to $t_f = 200$ days (the approximate number of school days in one academic year), and the interval $[0, t_f]$ was discretized into $n = 4{,}000$ equally spaced grid points, giving a time step $\Delta t = t_f/(n-1)$. The parameter set in Table 3 yields

$$\mathcal{R}_0 = \frac{\alpha\Lambda}{\mu(\theta+\mu)} = \frac{(8\times10^{-5})(0.315)}{(9\times10^{-4})(0.013+9\times10^{-4})} \approx 2.014,$$

confirming that the baseline scenario corresponds to an endemic bullying equilibrium ($\mathcal{R}_0 > 1$). The values of δ and σ follow Crokidakis [8], and the remaining parameters are grounded in the school setting: μ gives an average schooling duration of about 5.6 years, Λ fixes the school size $\Lambda/\mu \approx 350$, θ reflects the effectiveness of anti-bullying programs [12], and α, β lie in the small mass-action range used in prior bullying models [8, 3, 1]. As these values are illustrative rather than calibrated, the robustness check below confirms the conclusions are insensitive to them.

The initial conditions were $S(0) = 275$, $B(0) = 20$, $E(0) = 40$, $V(0) = 15$, and $R(0) = 0$, and the weight constants in the objective functional were $C_1 = 5$, $C_2 = 50$, $A_1 = 2$, and $A_2 = 1$. The weights reflect features of the program. Taking $A_1 > A_2$ places greater priority on reducing the number of active perpetrators than on protecting students who are merely at risk, while $C_2 \gg C_1$ reflects that broad victimization-prevention measures are more costly to sustain than targeted disciplinary action against identified perpetrators. Such measures include school-wide supervision, peer support, and anti-bullying initiatives.

Table 3: Parameter values used in the numerical simulation.

Parameter	Value	Source
Λ	0.3150	Assumed
α	0.00008	Assumed
β	0.00020	Assumed
θ	0.0130	Assumed
η	0.0055	Assumed
μ	0.0009	Assumed
δ	0.0500	[8]
σ	0.0700	[8]

We stress, however, that the ineffectiveness of the u_2-only strategy is not an artifact of this cost ratio; it is structural. It follows from the form of R_0, which depends on the perpetrator-generation rate α but not on the victimization rate β. Lowering C_2 therefore reduces the cost of deploying u_2 while leaving R_0, and hence its capacity to suppress the bulling dynamics, untouched.

Three intervention strategies were investigated:

- **Strategy 1** (perpetrator-focused only): $u_1 \neq 0$, $u_2 = 0$.
- **Strategy 2** (victimization-prevention only): $u_1 = 0$, $u_2 \neq 0$.
- **Strategy 3** (combined intervention): $u_1 \neq 0$, $u_2 \neq 0$.

The control profiles are presented in Figure 3. The perpetrator-focused control $u_1(t)$ remains at its upper bound for approximately the first 150 days before declining monotonically to zero at t_f, indicating that intensive early intervention is essential during the most critical phase of the school year. Moreover, the profiles of u_1 under Strategies 1 and 3 are virtually identical, confirming that the presence of u_2 does not alter the optimal timing of perpetrator-focused measures.

The victimization-prevention control $u_2(t)$ exhibits a markedly different profile. Under Strategy 2 (applied alone), u_2 rises gradually from zero, peaks at day 140, and then decreases toward t_f. Under Strategy 3 (combined), u_2 starts at a high level ($u_2(0) \approx 0.85$) and decreases monotonically throughout the horizon. The higher initial intensity of u_2 in Strategy 3 reflects the synergistic interaction between the two controls: when u_1 suppresses perpetrator numbers early, the marginal benefit of victimization-prevention is greatest at the start of the intervention.

The objective functional values are compared in Figure 4 and Table 4. In the absence of control, the objective attains $J_0 = 18{,}171.01$. Strategy 1 reduces J to 7,951.98, a decrease of

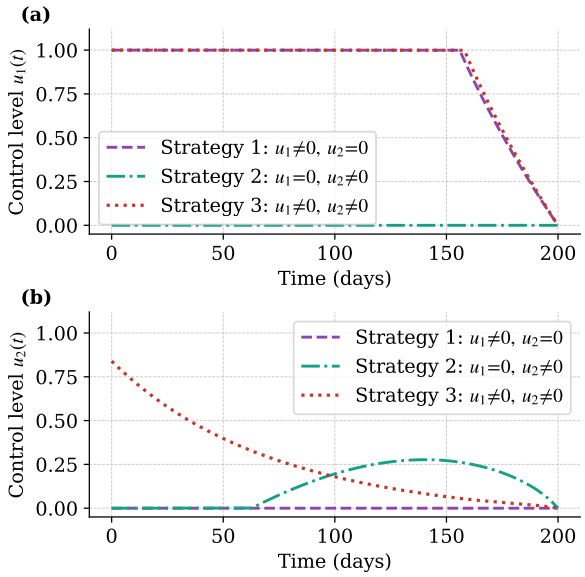


Figure 3: Optimal control profiles under the three intervention strategies: (a) perpetrator-focused control $u_1(t)$, and (b) victimization-prevention control $u_2(t)$. Strategy 2 sets $u_1 \equiv 0$ by construction; Strategy 1 sets $u_2 \equiv 0$.

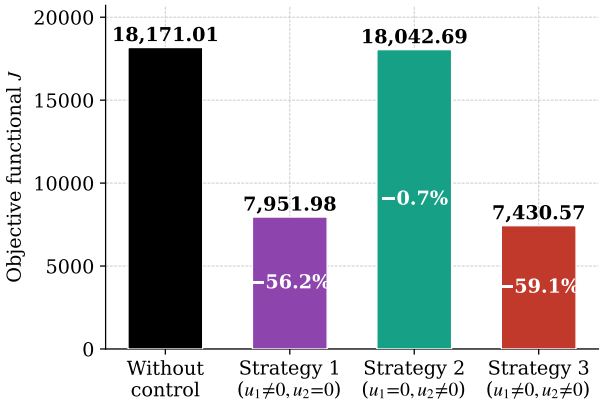


Figure 4: Comparison of the objective functional J across the uncontrolled baseline and the three intervention strategies. Percentage values inside each bar indicate the reduction relative to the uncontrolled value $J_0 = 18{,}171.01$.

approximately 56.24% relative to the baseline. The combined Strategy 3 achieves a further reduction to $J = 7{,}430.57$, corresponding to an overall decrease of 59.11%. By contrast, Strategy 2 yields a negligible improvement of only 0.71%, with $J = 18{,}042.69$, confirming that victimization prevention alone is insufficient to reduce the bullying burden in a cost-effective manner.

These results are consistent with the structure of the basic reproduction number

$$\mathcal{R}_0 = \frac{\alpha\Lambda}{\mu(\theta+\mu)},$$

which depends directly on the perpetrator-generation pathway through the parameter α but is independent of the victimization rate β. Because $\mathcal{R}_0$ governs the persistence of bullying in the population, any intervention that does not target the pathway controlled by α cannot substantially alter the endemic equilibrium. The simulated population trajectories are displayed in Figures 5–6, and the terminal values are summarized in Table 4.

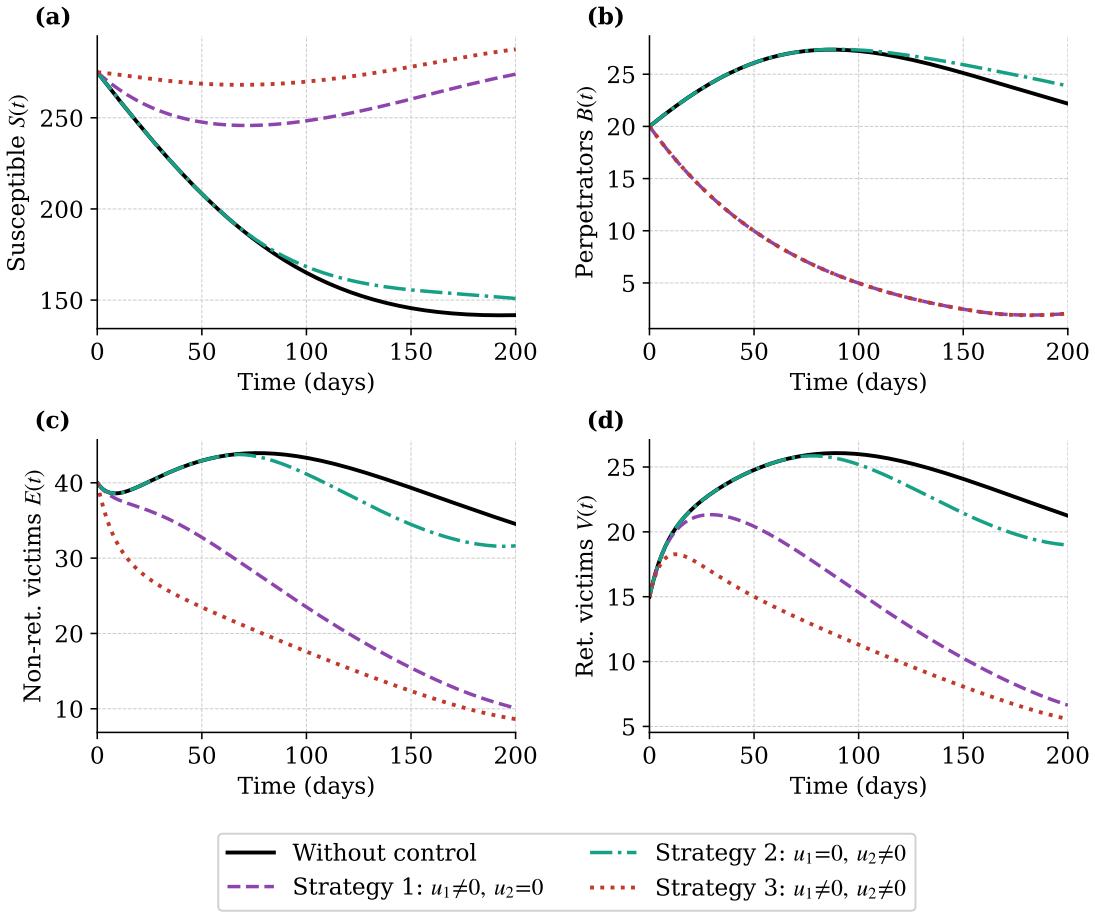


Figure 5: Simulated population trajectories under the three intervention strategies and the uncontrolled baseline: (a) susceptible students $S(t)$, (b) bullying perpetrators $B(t)$, (c) nonretaliating victims $E(t)$, and (d) retaliating victims $V(t)$.

Under the uncontrolled scenario and Strategy 2, the susceptible population declines steadily throughout the horizon as students are recruited into bullying-related compartments. Strategies 1 and 3, by suppressing perpetrator recruitment, preserve a substantially larger susceptible

Table 4: Terminal population values under each scenario. S_0: uncontrolled baseline; S_1, S_2, S_3: Strategies 1–3.

Compartment	S_0	S_1	S_2	S_3
$S(t_f)$	142	274	151	288
$B(t_f)$	22	2	24	2
$E(t_f)$	35	10	32	9
$V(t_f)$	21	7	19	6
$R(t_f)$	130	57	125	46
J	18,171.01	7,951.98	18,042.69	7,430.57
Reduction (%)	—	56.24	0.71	59.11

population: $S(t_f) = 274$ under Strategy 1 and $S(t_f) = 288$ under Strategy 3, compared with $S(t_f) = 142$ in the uncontrolled case.

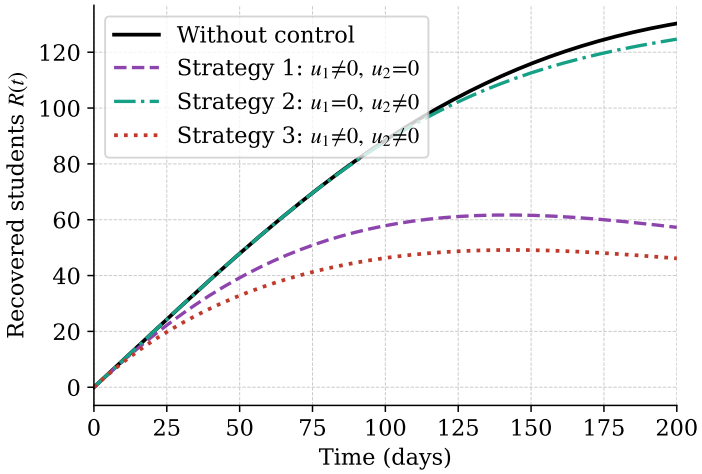


Figure 6: Recovered student population $R(t)$ under the three intervention strategies and the uncontrolled baseline. A large recovered population reflects greater cumulative exposure to bullying, not a favorable outcome.

The perpetrator population evolves most dramatically under the controls that include u_1. Under Strategies 1 and 3, $B(t)$ declines continuously from the initial value $B(0) = 20$ to terminal values of 2.065 and 2.070, respectively—reductions of more than 90%. Strategy 2 fails to reduce $B(t)$; indeed, the terminal perpetrator count (24) slightly *exceeds* the uncontrolled value (22), because diverting susceptibles away from victimization does not interrupt the self-reinforcing perpetrator-generation pathway [6].

Both victim compartments are reduced most effectively by Strategy 3. The terminal values $(E(t_f), V(t_f)) = (9, 6)$ under Strategy 3 represent reductions of 75.0% and 73.8%, respectively,

relative to the uncontrolled case. Strategy 1 achieves comparable but slightly smaller reductions ($E(t_f) = 10$, $V(t_f) = 7$), while Strategy 2 produces only marginal improvements.

Counterintuitively, the uncontrolled scenario yields the largest recovered population ($R(t_f) = 130$). This reflects the model structure: recovery occurs only after students have already passed through bullying-related compartments. A large $R(t_f)$ therefore indicates a heavier cumulative burden of bullying exposure, not a favorable epidemiological outcome. Strategies 1 and 3 produce the smallest recovered populations (57 and 46, respectively) because fewer students enter the bullying process in the first place.

Finally, in Table 5, varying each assumed parameter ($\Lambda, \alpha, \beta, \theta, \eta, \mu$) by $\pm 25\%$ and re-solving the optimality system leaves the ranking unchanged: Strategy 3 is always the best and Strategy 2 is always the worst. This reflects the structure of R_0, which depends on α but not β, rather than the precise parameter values.

Table 5: Robustness of the strategy ranking to $\pm 25\%$ variation in each assumed parameter. In every case Strategy 3 attains the lowest objective value and Strategy 2 the highest, so the ranking is preserved.

Parameter	Factor	J (S1)	J (S2)	J (S3)	Ranking preserved
Λ	0.75	7,914.1	17,598.2	7,402.4	yes
Λ	1.25	7,989.8	18,499.6	7,458.4	yes
α	0.75	7,940.3	14,445.5	7,419.5	yes
α	1.25	7,959.6	21,583.3	7,437.8	yes
β	0.75	7,359.4	17,963.0	7,046.3	yes
β	1.25	8,513.8	18,132.2	7,740.9	yes
θ	0.75	9,991.5	22,932.2	9,105.1	yes
θ	1.25	6,496.4	14,262.4	6,170.0	yes
η	0.75	7,934.4	17,719.8	7,418.9	yes
η	1.25	7,967.9	18,344.3	7,441.0	yes
μ	0.75	8,102.3	18,664.9	7,555.5	yes
μ	1.25	7,806.3	17,446.4	7,308.8	yes

6 Conclusion

This study developed and analyzed a compartmental mathematical model of school bullying dynamics incorporating a recovery compartment. The model was examined in two settings: without and with optimal control. For the uncontrolled model, the basic reproduction number $\mathcal{R}_0$ was derived as a threshold parameter governing the persistence of bullying. The bullying-free equilibrium was shown to be locally asymptotically stable when $\mathcal{R}_0 < 1$ and globally asymptotically stable under the same condition.

The model was extended by introducing two time-dependent control variables: $u_1(t)$, representing punishment-based intervention targeting perpetrator formation, and $u_2(t)$, representing victimization-prevention measures. The optimality system was derived via Pontryagin's Maximum Principle and solved numerically using a fourth-order Runge–Kutta scheme within the forward–backward sweep method.

Three intervention strategies were evaluated and compared against an uncontrolled baseline ($J_0 = 18{,}171.01$). Strategy 1 (perpetrator-focused only) reduced the objective functional by 56.24%, to $J = 7{,}951.98$, while Strategy 2 (victimization-prevention only) yielded a negligible reduction of 0.71%, with $J = 18{,}042.69$. Strategy 3 (combined intervention) achieved the greatest reduction of 59.11%, with $J = 7{,}430.57$. These results demonstrate that perpetrator-focused intervention is the primary driver of bullying suppression, consistent with the structure of $\mathcal{R}_0$, which depends on the perpetrator-generation rate α but not on the victimization rate β. The combined strategy, however, produced the lowest victim populations at the terminal time, confirming that supplementing perpetrator-focused measures with victimization-prevention support yields the most cost-effective outcome.

Several limitations suggest directions for future research. The model assumes a single recovery rate across the bullying-related classes; allowing class-specific recovery rates $\theta_B, \theta_E, \theta_V$ is a natural extension. In addition, calibrating the model against empirical bullying data would strengthen its applied relevance.

Funding

This research received no external funding.

Conflicts of Interest

The authors declares no conflicts of interest.

References

[1] Adeniji, A.A., Addai, E.A., Michael, S., Malesela, K.M., Asamoah, J.K.K. and Oshinubi, K. *Analysis of school bullying menace incorporating family education: A mathematical modelling approach*, Front. Appl. Math. Stat. 10 (2024), 1502500.

[2] Arseneault, L. *Annual research review: The persistent and pervasive impact of being bullied in childhood and adolescence: Implications for policy and practice*, J. Child Psychol. Psychiatry 59(4) (2018), 405–421.

[3] Ashi, H.A. *Stability analysis of a simple mathematical model for school bullying*, AIMS Math. 7(4) (2022), 4936–4945.

[4] Castillo-Chavez, C., Feng, Z. and Huang, W. *On the computation of $\mathcal{R}_0$ and its role on global stability*, in *Mathematical approaches for emerging and re-emerging infectious diseases: An introduction*, IMA Vol. Math. Appl. 125 (2002), 229–250.

[5] Castillo-Chavez, C. and Song, B. *Dynamical models of tuberculosis and their applications*, Math. Biosci. Eng. 1(2) (2004), 361–404.

[6] Christakis, N.A. and Fowler, J.H. *Social contagion theory: Examining dynamic social networks and human behavior*, Stat. Med. 32(4) (2013), 556–577.

[7] Copeland, W.E., Wolke, D., Angold, A. and Costello, E.J. *Adult psychiatric outcomes of bullying and being bullied by peers in childhood and adolescence*, JAMA Psychiatry 70(4) (2013), 419–426.

[8] Crokidakis, N. *A mathematical model for the bullying dynamics in schools*, Appl. Math. Comput. 492 (2025), 129254.

[9] Espelage, D.L. and Swearer, S.M. *A social-ecological model for bullying prevention and intervention: Understanding the impact of adults in the social ecology of youngsters*, in *Handbook of bullying in schools*, Routledge, (2010), 61–72.

[10] Farrington, D.P. and Ttofi, M.M. *School-based programs to reduce bullying and victimisation*, Campbell Syst. Rev. 5(1) (2009), 1–148.

[11] Fleming, W.H. and Rishel, R.W. *Deterministic and stochastic optimal control*, Applications of Mathematics, Vol. 1, Springer-Verlag, New York, 1975.

[12] Gaffney, H., Farrington, D.P. and Ttofi, M.M. *Examining the effectiveness of school-bullying intervention programs globally: A meta-analysis*, Int. J. Bullying Prev. 3 (2021), 14–31.

[13] Huitsing, G. and Veenstra, R. *Bullying in classrooms: Participant roles from a social network perspective*, Aggress. Behav. 38(6) (2012), 494–509.

[14] Id Ouaziz, S. and El Khomssi, M. *Exploring corruption through mathematical modeling and optimal control analysis: Examining interactions between students and employees*, J. Math. 2025 (2025), Article ID 2621856.

[15] Izzati, N. and Ningsih, W. *Rehabilitation and law enforcement as optimal controls in a mathematical model of social behavior*, J. Statistika 18 (2025), 920–929.

[16] Lenhart, S. and Workman, J.T. *Optimal control applied to biological models*, Chapman & Hall/CRC, 2007.

[17] Murray, J.D. *Mathematical biology I: An introduction*, 3rd ed., Interdisciplinary Applied Mathematics, Vol. 17, Springer, New York, 2002.

[18] Nowzari, C., Preciado, V.M. and Pappas, G.J. *Analysis and control of epidemics: A survey of spreading processes on complex networks*, IEEE Control Syst. Mag. 36(1) (2016), 26–46.

[19] Perko, L. *Differential equations and dynamical systems*, 3rd ed., Texts in Applied Mathematics 7, Springer, New York, 2001.

[20] Pontryagin, L.S., Boltyanskii, V.G., Gamkrelidze, R.V. and Mishchenko, E.F. *The mathematical theory of optimal processes*, Interscience Publishers, New York, 1962.

[21] Salmivalli, C. *Bullying and the peer group: A review*, Aggress. Violent Behav. 15(2) (2010), 112–120.

[22] Teklu, S.W. and Terefe, B.B. *Mathematical modeling analysis on the dynamics of university students' animosity towards mathematics with optimal control theory*, Sci. Rep. 12 (2022), 11578.

[23] Ttofi, M.M., Farrington, D.P., Lösel, F. and Loeber, R. *The predictive efficiency of school bullying versus later offending: A systematic/meta-analytic review of longitudinal studies*, Crim. Behav. Ment. Health 21(2) (2011), 80–89.

[24] van den Driessche, P. and Watmough, J. *Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission*, Math. Biosci. 180 (2002), 29–48.

[25] Wolke, D. and Lereya, S.T. *Long-term effects of bullying*, Arch. Dis. Child. 100(9) (2015), 879–885.

Iranian Journal of Numerical Analysis and Optimization
Vol. 17, No. 1, 2027, pp 177–216
https://doi.org/10.22067/ijnao.2026.98719.1868
https://ijnao.um.ac.ir/

CC BY NC

Research Article

Modeling the transmission dynamics of meningococcal meningitis under hospitalization effects

I.O. Joshua, K. Kiruthika, M. Gümüş, A. Raza, E. Chandrakumar and N. Jeeva*,

Abstract

Meningococcal meningitis remains a major public health concern owing to its rapid transmission, severe health consequences, and potential to cause recurrent outbreaks. In this study,

*Corresponding author

Received 13 May 2026; revised 31 July 2026; accepted 1 August 2026

Idowu Oluwarimi Joshua
Department of Mathematics, University of Benin, Edo State, Nigeria. e-mail: joshua.idowu@physci.uniben.edu
K. Kiruthika
Department of Mathematics, K.S.Rangasamy College of Technology, Tiruchengode 637215, Tamil Nadu, India. e-mail: kiruthika@ksrct.ac.in
Mehmet Gümüş
Department of Mathematics, Faculty of Science, Zonguldak Bülent Ecevit University, 67100, Zonguldak, Türkiye. e-mail: m.gumus@beun.edu.tr
Ali Raza
Department of Mathematics, School of Science, University of Management and Technology, Lahore-54770, Pakistan. e-mail: alig03645@gmail.com
E. Chandrakumar
Department of Electrical and Electronics Engineering, K.S.Rangasamy College of Technology, Tiruchengode 637215, Tamil Nadu, India. e-mail: chandrakumar@ksrct.ac.in
N. Jeeva
Department of Mathematics, K.S.Rangasamy College of Technology, Tiruchengode 637215, Tamil Nadu, India. e-mail: jeevapirc2405@gmail.com

How to cite this article
Joshua, I.O., Kiruthika, K., Gümüş, M., Raza, A., Chandrakumar, E. and Jeeva, N., Modeling the transmission dynamics of meningococcal meningitis under hospitalization effects. *Iran. J. Numer. Anal. Optim.*, 2027; 17(1): 177-216. https://doi.org/10.22067/ijnao.2026.98719.1868

a deterministic compartmental model is developed to investigate the transmission dynamics of meningococcal meningitis and assess the impact of hospitalization on disease spread. The total population is partitioned into five epidemiological classes: susceptible (S), exposed (E), infectious (I), hospitalized (H), and recovered (R). The proposed model is shown to be epidemiologically and mathematically well-posed based on the positivity and boundedness of its solutions. The disease-free and endemic equilibrium points are derived, and the basic reproduction number, $\mathcal{R}_0$, is obtained using the next-generation matrix method. It is proved that the disease-free equilibrium is locally and globally asymptotically stable whenever $\mathcal{R}_0 < 1$, whereas the unique endemic equilibrium is locally and globally asymptotically stable whenever $\mathcal{R}_0 > 1$. Furthermore, bifurcation analysis based on the center manifold theory of Castillo–Chavez and Song establishes that the model undergoes a forward (supercritical) bifurcation at the critical threshold $\mathcal{R}_0 = 1$, implying that reducing the basic reproduction number below unity is sufficient for disease eradication. Sensitivity analysis identified the transmission, hospitalization, and recovery rates as the most influential parameters governing disease transmission. Numerical simulations using MATLAB® supported the analytical results and demonstrated that timely hospitalization, improved recovery, and reduced transmission substantially decreased the number of infectious individuals and shortened the outbreak duration. These findings provide quantitative insights into meningococcal meningitis transmission and offer guidance for designing effective interventions and disease control strategies.

AMS subject classifications (2020): Primary 26A33; Secondary 65L05, 92D30.

Keywords: Meningitis modeling; Treatment effects; Epidemic model; Stability analysis; Numerical simulation

1 Introduction

Meningococcal infection has been acknowledged as a severe issue for nearly 200 years. When an outbreak raced across Geneva in 1805, Vieusseux described meningococcal meningitis for the first time [23]. In 1887, Neisseria meningitidis was identified as the causative agent of meningococcal disease. Approximately 10% of people in the general population carry meningococcal germs in their nose and throat without experiencing any adverse consequences. Meningococci can occasionally evade the body's defenses, resulting in life-threatening illnesses. It causes fulminant meningococcemia and meningococcal meningitis, two potentially fatal illnesses that frequently coexist. Neisseria meningitidis continues to be a major cause of meningitis and deadly sepsis, despite the availability of effective medicines and partially successful vaccinations. Meningitis is an inflammatory reaction to bacterial infection of the subarachnoid space and leptomeninges (pia-arachnoid). Meningococcemia with petechiae can be caused by bacteria growing in the

bloodstream or seeding local sites such as the pericardium, joints, or meninges. In as many as 50% of instances, it can result in multi-organ failure, severe disability, or death [16, 27]. Adults with fever and maculopapular rash have a wide range of potential diagnoses. The differential diagnosis encompasses viral and bacterial infections and hypersensitivity reactions, such as those triggered by drugs. Some of these eruptions constitute medical emergencies, whereas others are benign [22].

One of the riskiest side effects of meningococcal meningitis is Waterhouse–Friderichsen syndrome. It is believed to be a coagulopathy that causes hemorrhagic necrosis of the adrenal glands, triggered by bacterial endotoxins. Petechial hemorrhages in the skin and mucous membranes are the defining features of this condition. Purpuric rashes can be caused by immune-mediated disorders, platelet dysfunction, vascular injury, or disseminated intravascular coagulation [13]. The fatality rate from meningococcal meningitis ranges from 5% to 15%, and if treatment is not received, it may increase to 50%. Nine percent of patients with invasive meningococcal illness have long-lasting sequelae, and more than one-third of patients survive but face significant problems [7, 8, 24].

Few studies have focused on creating mathematical models to examine the spread of meningococcal meningitis. They are all based on differential equations and have some limitations, such as not accounting for the unique characteristics of populations, not simulating the behavior of each individual, and not being suitable for computational implementation. Mathematical models of the dynamics of disease transmission are essential for better understanding disease behavior, optimizing the use of scarce resources, and suggesting control methods for infectious diseases. Decisions regarding intervention techniques to prevent and control the resurgence of meningitis and pneumonia are more successful when guided by models specifically designed for these diseases. Given the emergence and resurgence of diseases, mathematical modeling is an efficient and appealing field for estimating the effects of control measures and the underlying parameters of real-world phenomena that are difficult or costly to obtain through experimentation [11, 9, 20, 19].

Mathematical modeling serves as a vital framework for analyzing the spread and control of bacterial meningitis within a population, enabling researchers to understand transmission pathways and evaluate intervention measures such as vaccination and contact reduction [28, 1, 4, 25, 14, 12, 15, 35]. These models provide insights into the dynamics of Neisseria meningitidis infections, helping to predict epidemic patterns, assess the impact of immunization programs such as the MACV vaccine, and inform public health decision-making [34, 21]. Furthermore, optimal control and sensitivity analyses within these models support the identification of effective strategies to minimize infection rates and mitigate the disease burden in the affected regions [2, 32]. In this study, Anderson and May [3] analyzed the relationship between infection and human

populations using epidemiological data and mathematical models to support the implementation of public-health interventions. Karachaliou et al. [17] developed a mathematical model of MenA transmission, predicting effective short-term control after the introduction of MenAfriVac but a possible resurgence without booster vaccinations. In addition, Keeling and Rohani [18] presented comprehensive infectious disease models to predict, understand, and manage outbreaks in humans and animals using actual data and simulation tools.

Furthermore, Trotter [29] reported that the predominance of Neisseria meningitidis serogroups varies geographically, with conjugate vaccines, especially for serogroup C, significantly reducing the disease incidence in developed nations. They further suggested that global control could be achieved through serogroup B vaccines in developed countries and serogroup A conjugates in Africa. Vereen [31] developed a meningitis model incorporating a carrier class into the SIR framework to assess vaccination effects in epidemic-prone regions. Their analysis demonstrated that higher transmission rates require proportionally higher vaccination rates, as supported by MATLAB simulations using WHO parameters. In addition, the WHO [33] highlighted that Neisseria meningitidis serogroups A, B, and C cause most meningococcal cases, with increasing outbreaks of Y and W135. Vaccination using conjugate or polysaccharide vaccines targeting specific serogroups remains the most effective approach for preventing and controlling these epidemics. Recently, [30] identified four major demographic trends shaping the world: population growth, aging, urbanization, and international migration. The report projects that the global population will likely peak before 2100 owing to declining birth rates, emphasizing how these demographic shifts affect sustainability, gender equality, and poverty reduction, and offering policy recommendations to help nations prepare for future population changes.

If not treated promptly, meningococcal meningitis, an acute bacterial infection affecting the meninges, can rapidly lead to severe complications, hospitalization, and death. The disease spreads mainly through close contact and respiratory droplets, with both symptomatic and pre-symptomatic individuals contributing to the transmission. In this study, we develop a novel mathematical model to characterize the transmission dynamics of meningococcal meningitis. The remainder of this paper is organized as follows: Section 2 presents the model formulation, description, and corresponding system of ODEs. Section 3 discusses the basic qualitative analysis of the proposed model. Section 4 focuses on the key theoretical aspects, including the computation of the basic reproduction number ($\mathcal{R}_0$) and identification of disease-free and endemic equilibrium points. Section 5 examines the stability analysis and backward bifurcation, and Section 6 explores the sensitivity analysis of the proposed model. Section 7 provides the numerical simulations and discussion of the results, and Section 8 concludes the study with future research directions.

2 Formulation of the meningococcal meningitis model with hospitalization

In this section, a deterministic compartmental model is formulated to describe the transmission dynamics of meningococcal meningitis in the presence of hospitalization. The model consists of a system of nonlinear ordinary differential equations that captures the progression of individuals through different epidemiological stages and examines the effects of hospitalization, recovery, disease-induced mortality, and waning immunity on disease transmission.

The total living population at time t, denoted by $N(t)$, is partitioned into five mutually exclusive epidemiological classes: susceptible $S(t)$, exposed $E(t)$, infectious $I(t)$, hospitalized $H(t)$, and recovered $R(t)$. Disease-induced deaths are incorporated through the mortality rates of infectious and hospitalized individuals and were therefore not represented as a separate compartment.

The model is developed under the following assumptions:

- The susceptible population is replenished through logistic recruitment with intrinsic recruitment rate b and carrying capacity K.
- Disease transmission occurs through effective contact between susceptible individuals and both exposed and infectious individuals. The exposed class contributes to disease transmission at a reduced level, measured by the modification parameter σ_1.
- Exposed individuals progress to the infectious class at a constant rate θ.
- Infectious individuals may be hospitalized at rate α, recover naturally at rate γ_1, or die due to disease-induced complications at rate δ_1.
- Hospitalized individuals either recover following treatment at rate γ_2 or die due to disease-induced complications at rate δ_2.
- Recovered individuals lose temporary immunity at rate ψ and return to the susceptible class, thereby allowing reinfection.
- All living compartments experience natural mortality at the constant rate μ.

2.1 Description of the model

The susceptible population increases through logistic recruitment represented by

$$bN\left(1-\frac{N}{K}\right),$$

where b denotes the intrinsic recruitment rate, and K is the carrying capacity of the population. Susceptible individuals acquire infection through effective contact with infectious and exposed individuals at the force of infection

$$\lambda = \beta\frac{I+\sigma_1 E}{N},$$

where β denotes the effective transmission coefficient, and σ_1 measures the relative infectiousness of exposed individuals compared to infectious individuals. Following infection, exposed individuals progress to the infectious class at the rate θ. Infectious individuals may be hospitalized at rate α, recover naturally at rate γ_1, or die due to disease-induced complications at rate δ_1. Hospitalized individuals recover under medical care at a rate γ_2 or die from disease-induced complications at a rate δ_2. Individuals in the recovered class lose temporary immunity at the rate ψ and re-enter the susceptible class, thereby contributing to reinfection. In addition, all living compartments are subject to natural mortality at a constant rate μ. The interactions among the epidemiological classes are illustrated in the compartmental flow diagram in Figure 1.

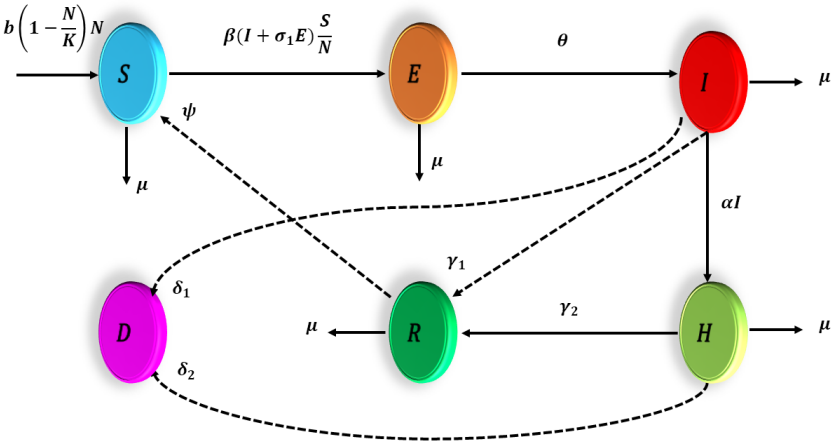


Figure 1: Transmission dynamics of the meningococcal meningitis model

2.2 Model equations

Given these assumptions, the mathematical formulation of the meningococcal meningitis model with hospitalization is presented below:

$$\begin{aligned}
\frac{dS}{dt} &= bN\left(1-\frac{N}{K}\right)-\beta(I+\sigma_1 E)\frac{S}{N}-\mu S+\psi R,\\
\frac{dE}{dt} &= \beta(I+\sigma_1 E)\frac{S}{N}-(\theta+\mu)E,\\
\frac{dI}{dt} &= \theta E-(\delta_1+\mu+\gamma_1+\alpha)I,\\
\frac{dH}{dt} &= \alpha I-(\mu+\delta_2+\gamma_2)H,\\
\frac{dR}{dt} &= \gamma_1 I+\gamma_2 H-(\mu+\psi)R,\\
\frac{dD}{dt} &= \delta_1 I+\delta_2 H,
\end{aligned} \tag{1}$$

where $N(t)=S(t)+E(t)+I(t)+H(t)+R(t)$ denotes the total population.
The corresponding initial conditions are

$$S(0)>0,\quad E(0)\geq 0,\quad I(0)\geq 0,\quad H(0)\geq 0,\quad R(0)\geq 0,\quad D(0)\geq 0.$$

Table 1 summarizes the compartments utilized in the proposed model (1), and Table 2 lists the model parameters, along with their corresponding descriptions and numerical values.

Table 1: Meningococcal meningitis model variables

Variable	Description
$S(t)$	Population of susceptible individuals
$E(t)$	Population of exposed individuals
$I(t)$	Population of infectious individuals
$H(t)$	Population of hospitalized individuals
$R(t)$	Population of recovered individuals
$D(t)$	Population of deceased individuals

Table 2: Proposed model parameters, descriptions, and their numerical values

Parameter	Description	Numerical Value	Source
b	Intrinsic per capita recruitment rate	0.02–0.05	[30]
K	Carrying capacity of the system	Normalized to 1	Assumed
μ	Natural death rate	0.000039	[29]
β	Transmission rate of the disease	0.3–1.5	[31]
σ_1	Meningitis exposed coefficient	0.2–0.85	[4]
θ	Rate of progression from exposed to infectious	0.1–0.52	[29]
γ_1	Recovery rate of infected individuals	0.05–0.2	[3]
γ_2	Recovery rate of hospitalized individuals	0.03–0.1	[18]
δ_1	Induced death rate of infected individuals	0.005–0.05	[33]
δ_2	Induced death rate of hospitalized individuals	0.01–0.10	[33]
α	Hospitalization or isolation rate	0–1	Assumed
ψ	Rate of loss of immunity ($R \to S$)	0.04–2	[17]

3 Qualitative analysis

3.1 Nonnegativity of solutions

Theorem 1. Let

$$(S(0), E(0), I(0), H(0), R(0), D(0)) \in \mathbb{R}^6_+,$$

where $\mathbb{R}^6_+ = \{(S, E, I, H, R, D) : S, E, I, H, R, D \geq 0\}$. Then, the solutions of system (1) remain nonnegative for all $t \geq 0$. In other words, the nonnegative orthant $\mathbb{R}^6_+$ is positively invariant under the flow generated by system (1).

Proof. To establish positivity, it suffices to show that the vector field associated with system (1) points inward on each boundary hyperplane of the nonnegative orthant.
For $S = 0$, we have

$$\left.\frac{dS}{dt}\right|_{S=0} = bN\left(1 - \frac{N}{K}\right) + \psi R \geq 0.$$

Similarly, for $E = 0$, $I = 0$, $H = 0$, $R = 0$, $D = 0$, it follows that

$$\left.\frac{dE}{dt}\right|_{E=0} = \beta\frac{SI}{N} \geq 0,$$
$$\left.\frac{dI}{dt}\right|_{I=0} = \theta E \geq 0,$$
$$\left.\frac{dH}{dt}\right|_{H=0} = \alpha I \geq 0,$$
$$\left.\frac{dR}{dt}\right|_{R=0} = \gamma_1 I + \gamma_2 H \geq 0,$$
$$\left.\frac{dD}{dt}\right|_{D=0} = \delta_1 I + \delta_2 H \geq 0.$$

Thus, on every boundary component of $\mathbb{R}^6_+$, the corresponding derivative is nonnegative. Hence, the vector field does not point outward across any coordinate hyperplane. By the forward invariance theorem (also known as Nagumo's Theorem), the nonnegative orthant $\mathbb{R}^6_+$ is positively invariant. Therefore, any solution of system (1) that starts in $\mathbb{R}^6_+$ remains in $\mathbb{R}^6_+$ for all $t \geq 0$. Consequently, $S(t), E(t), I(t), H(t), R(t), D(t) \geq 0$, for $t \geq 0$. □

3.2 Boundedness

Theorem 2. The solutions of system (1) are bounded in the feasible region

$$\Omega = \left\{(S, E, I, H, R, D) \in \mathbb{R}^6_+ : N(t) \leq \frac{K(b-\mu)}{b}\right\}.$$

Proof. Let $N(t) = S(t)+E(t)+I(t)+H(t)+R(t)$ denotes the total living population. Summing the first five equations of system (1) yields

$$\frac{dN}{dt} = bN\left(1 - \frac{N}{K}\right) - \mu N - \delta_1 I - \delta_2 H.$$

Since $\delta_1 I + \delta_2 H \geq 0$, it follows that

$$\frac{dN}{dt} \leq bN\left(1 - \frac{N}{K}\right) - \mu N.$$

Hence

$$\frac{dN}{dt} \leq (b-\mu)N - \frac{b}{K}N^2.$$

Consider the auxiliary logistic equation

$$\frac{dy}{dt} = (b-\mu)y - \frac{b}{K}y^2, \quad y(0) = N(0).$$

Its positive equilibrium is $y^* = \frac{K(b-\mu)}{b}$. By the comparison theorem, $N(t) \leq y(t), \quad t \geq 0$. Since every solution of the logistic equation is bounded above by its carrying capacity y^*, we obtain

$$N(t) \leq \frac{K(b-\mu)}{b}, \qquad t \to \infty.$$

Consequently,

$$0 \leq N(t) \leq \max\left\{N(0), \frac{K(b-\mu)}{b}\right\}, \qquad t \geq 0.$$

Therefore, all state variables $S(t), E(t), I(t), H(t), R(t)$ are bounded for all $t \geq 0$, and the region

$$\Omega = \left\{(S, E, I, H, R, D) \in \mathbb{R}^6_+ : N(t) \leq \frac{K(b-\mu)}{b}\right\},$$

is positively invariant. Hence, the solutions of system (1) remain bounded for all times. □

4 Threshold analysis of the model

In this section, the threshold dynamics of the proposed model are investigated by determining the disease-free equilibrium (DFE), basic reproduction number, and endemic equilibrium. The conditions for the existence and stability of these equilibria are subsequently established.

4.1 Disease-free equilibrium

The DFE of system (1) corresponds to the steady state at which meningococcal meningitis is absent from the population. Setting $E = I = H = R = 0$ in system (1), the susceptible equation becomes

$$0 = bN\left(1 - \frac{N}{K}\right) - \mu S.$$

Since $N = S$ at the DFE, we have

$$0 = bS\left(1 - \frac{S}{K}\right) - \mu S.$$

Factoring out S yields

$$S\left[b\left(1-\frac{S}{K}\right)-\mu\right]=0.$$

The trivial solution $S=0$ is thus biologically irrelevant. Hence, the biologically feasible DFE is given by

$$S^*=K\left(1-\frac{\mu}{b}\right),\quad b>\mu.$$

Therefore, the DFE of system (1) is

$$\mathcal{E}_0=\left(K\left(1-\frac{\mu}{b}\right),0,0,0,0\right).$$

4.2 Basic reproduction number

The basic reproduction number, denoted by $\mathcal{R}_0$, is defined as the average number of secondary infections generated by a single infectious individual introduced into a completely susceptible population. Following the next-generation matrix approach of [26, 10, 5], we consider the infected-state vector

$$\mathbf{x}=(E,I)^T.$$

The vectors of new infection and transition terms are

$$F(\mathbf{x})=\begin{pmatrix}\beta(\sigma_1E+I)\\0\end{pmatrix},\quad V(\mathbf{x})=\begin{pmatrix}(\theta+\mu)E\\-\theta E+dI\end{pmatrix},$$

where $d=\alpha+\delta_1+\gamma_1+\mu$. The corresponding Jacobian matrices evaluated at the DFE are

$$F=\begin{pmatrix}\beta\sigma_1 & \beta\\0 & 0\end{pmatrix},\quad V=\begin{pmatrix}\theta+\mu & 0\\-\theta & d\end{pmatrix}.$$

The inverse of V is $V^{-1}=\frac{1}{(\theta+\mu)d}\begin{pmatrix}d & 0\\\theta & \theta+\mu\end{pmatrix}$. Hence,

$$FV^{-1}=\frac{1}{(\theta+\mu)d}\begin{pmatrix}\beta(\sigma_1d+\theta) & \beta(\theta+\mu)\\0 & 0\end{pmatrix}.$$

Therefore, the basic reproduction number is the spectral radius of the next-generation matrix, that is,

$$\begin{aligned}\mathcal{R}_0 &= \rho(FV^{-1}) \\ &= \frac{\beta\left[\theta + \sigma_1(\alpha + \delta_1 + \gamma_1 + \mu)\right]}{(\theta + \mu)(\alpha + \delta_1 + \gamma_1 + \mu)}.\end{aligned} \tag{2}$$

The threshold quantity $\mathcal{R}_0$ determines the disease transmission dynamics. In particular, meningococcal meningitis can invade the population if $\mathcal{R}_0 > 1$, whereas the disease eventually dies out when $\mathcal{R}_0 < 1$.

4.3 Existence of the endemic equilibrium

Theorem 3. System (1) admits a unique endemic equilibrium

$$\mathcal{E}^{**} = (S^{**}, E^{**}, I^{**}, H^{**}, R^{**}),$$

if and only if $\mathcal{R}_0 > 1$.

Proof. Let $\mathcal{E}^{**} = (S^{**}, E^{**}, I^{**}, H^{**}, R^{**})$ denote an arbitrary endemic equilibrium of system (1), where $N^{**} = S^{**} + E^{**} + I^{**} + H^{**} + R^{**}$.
Define the equilibrium force of infection by

$$\lambda^{**} = \frac{\beta(I^{**} + \sigma_1 E^{**})}{N^{**}}. \tag{3}$$

Setting the right-hand side of system (1) equal to zero and solving successively yields

$$\begin{aligned}E^{**} &= \frac{\lambda^{**} S^{**}}{k_1}, \\ I^{**} &= \frac{\theta\lambda^{**} S^{**}}{k_1 d}, \\ H^{**} &= \frac{\alpha\theta\lambda^{**} S^{**}}{k_1 d k_3}, \\ R^{**} &= \left(\frac{\gamma_1\theta}{k_1 d k_4} + \frac{\alpha\theta\gamma_2}{k_1 d k_3 k_4}\right)\lambda^{**} S^{**},\end{aligned} \tag{4}$$

where $k_1 = \theta + \mu$, $d = \alpha + \delta_1 + \gamma_1 + \mu$, $k_3 = \delta_2 + \gamma_2 + \mu$, $k_4 = \mu + \psi$. Substituting (4) into $N^{**} = S^{**} + E^{**} + I^{**} + H^{**} + R^{**}$, gives

$$N^{**} = S^{**}\left(1 + L_2\lambda^{**}\right), \tag{5}$$

where $L_2 = \frac{1}{k_1} + \frac{\theta}{k_1 d} + \frac{\alpha\theta}{k_1 d k_3} + \frac{\gamma_1\theta}{k_1 d k_4} + \frac{\alpha\theta\gamma_2}{k_1 d k_3 k_4}$. Substituting (4) and (5) into the force of infection (3) yields

$$\lambda^{**} = \frac{\beta(\theta + d\sigma_1)}{k_1 d} \times \frac{\lambda^{**}}{1 + L_2\lambda^{**}}. \tag{6}$$

Since $\lambda^{**} > 0$, (6) becomes

$$1 + L_2\lambda^{**} = \frac{\beta(\theta + d\sigma_1)}{k_1 d} = \mathcal{R}_0. \tag{7}$$

Hence,

$$\lambda^{**} = \frac{\mathcal{R}_0 - 1}{L_2}. \tag{8}$$

Because every parameter of the model is positive, it follows that $L_2 > 0$. Therefore,

$$\lambda^{**} > 0 \quad \Longleftrightarrow \quad \mathcal{R}_0 > 1.$$

Thus, system (1) possesses a unique biologically meaningful endemic equilibrium if and only if $\mathcal{R}_0 > 1$. □

Remark 1. The above theorem establishes the existence and uniqueness of the endemic equilibrium. However, it does not determine the nature of the bifurcation occurring at the threshold $\mathcal{R}_0 = 1$. This is investigated in the subsequent section using the center manifold theory of Castillo-Chavez and Song.

5 Stability analysis

In this section, the local and global stability properties of the proposed model's disease-free and endemic equilibria are investigated.

5.1 Local stability of the disease-free equilibrium

Theorem 4. The DFE $\mathcal{E}_0$ of system (1) is locally asymptotically stable whenever $\mathcal{R}_0 < 1$ and unstable whenever $\mathcal{R}_0 > 1$.

Proof. The Jacobian matrix of system (1) evaluated at the DFE $\mathcal{E}_0$ is

$$J(\mathcal{E}_0) = \begin{pmatrix} \mu - b & -\beta\sigma_1 & -\beta & 0 & \psi \\ 0 & \beta\sigma_1 - k_1 & \beta & 0 & 0 \\ 0 & \theta & -d & 0 & 0 \\ 0 & 0 & \alpha & -k_3 & 0 \\ 0 & 0 & \gamma_1 & \gamma_2 & -k_4 \end{pmatrix},$$

where $k_1 = \theta + \mu, \quad k_3 = \mu + \delta_2 + \gamma_2, \quad k_4 = \mu + \psi, \quad d = \alpha + \delta_1 + \gamma_1 + \mu$. The characteristic equation is obtained from $\det(\lambda I - J(\mathcal{E}_0)) = 0$, that is,

$$\begin{vmatrix} \lambda + b - \mu & \beta\sigma_1 & \beta & 0 & -\psi \\ 0 & \lambda - \beta\sigma_1 + k_1 & -\beta & 0 & 0 \\ 0 & -\theta & \lambda + d & 0 & 0 \\ 0 & 0 & -\alpha & \lambda + k_3 & 0 \\ 0 & 0 & -\gamma_1 & -\gamma_2 & \lambda + k_4 \end{vmatrix} = 0.$$

Since the first column contains only one nonzero entry, expanding along the first column gives

$$(\lambda + b - \mu) \begin{vmatrix} \lambda - \beta\sigma_1 + k_1 & -\beta & 0 & 0 \\ -\theta & \lambda + d & 0 & 0 \\ 0 & -\alpha & \lambda + k_3 & 0 \\ 0 & -\gamma_1 & -\gamma_2 & \lambda + k_4 \end{vmatrix} = 0.$$

The remaining determinant is a block lower triangular. Hence,

$$(\lambda + b - \mu) \begin{vmatrix} \lambda - \beta\sigma_1 + k_1 & -\beta \\ -\theta & \lambda + d \end{vmatrix} \begin{vmatrix} \lambda + k_3 & 0 \\ -\gamma_2 & \lambda + k_4 \end{vmatrix} = 0.$$

Therefore, $(\lambda + b - \mu)(\lambda + k_3)(\lambda + k_4)\left[(\lambda - \beta\sigma_1 + k_1)(\lambda + d) - \beta\theta\right] = 0$. Consequently, three eigenvalues are obtained immediately as

$$\lambda_1 = \mu - b, \quad \lambda_2 = -k_3, \quad \lambda_3 = -k_4.$$

Since $b > \mu, \quad k_3 > 0, \quad k_4 > 0$, it follows that $\lambda_1, \lambda_2, \lambda_3 < 0$. The remaining two eigenvalues satisfy

$$(\lambda - \beta\sigma_1 + k_1)(\lambda + d) - \beta\theta = 0,$$

or equivalently,

$$\lambda^2 + (k_1 + d - \beta\sigma_1)\lambda + (k_1 - \beta\sigma_1)d - \beta\theta = 0.$$

Using the definition of basic reproduction number $\mathcal{R}_0$, the constant term becomes $(k_1-\beta\sigma_1)d-\beta\theta = k_1 d(1-\mathcal{R}_0)$. Hence, the quadratic equation reduces to

$$\lambda^2 + (k_1 + d - \beta\sigma_1)\lambda + k_1 d(1-\mathcal{R}_0) = 0.$$

Since $k_1 + d - \beta\sigma_1 > 0$ and $k_1 d(1-\mathcal{R}_0) > 0$, whenever $\mathcal{R}_0 < 1$, the Routh–Hurwitz criterion guarantees that both roots have negative real parts. Therefore, all eigenvalues of the Jacobian have negative real parts for $\mathcal{R}_0 < 1$. Therefore, the DFE is locally asymptotically stable.

Conversely, if $\mathcal{R}_0 > 1$, then the constant term of the quadratic becomes negative, implying that one eigenvalue is positive. Hence, the DFE is unstable. This completes the proof. □

Theorem 5. The DFE of system (1) is globally asymptotically stable whenever $\mathcal{R}_0 < 1$.

Proof. Consider the Lyapunov function

$$V(E,I) = E + \frac{\beta}{d}I, \tag{9}$$

which is positive definite in the feasible region and vanishes only at the DFE.

Differentiating V along the solutions of system (1) gives

$$\begin{aligned}\dot{V} &= \dot{E} + \frac{\beta}{d}\dot{I} \\ &= \left[\beta(I+\sigma_1 E)\frac{S}{N} - (\theta+\mu)E\right] + \frac{\beta}{d}\left[\theta E - (\alpha+\delta_1+\gamma_1+\mu)I\right].\end{aligned}$$

Since $d = \alpha+\delta_1+\gamma_1+\mu$, we obtain

$$\dot{V} = \beta(I+\sigma_1 E)\frac{S}{N} - (\theta+\mu)E + \frac{\beta\theta}{d}E - \beta I.$$

Using the fact that $S \le N$, it follows that

$$\beta(I+\sigma_1 E)\frac{S}{N} \le \beta(I+\sigma_1 E),$$

and therefore

$$\begin{aligned}\dot{V} &\le \beta(I+\sigma_1 E) - (\theta+\mu)E + \frac{\beta\theta}{d}E - \beta I \\ &= \left(\beta\sigma_1 + \frac{\beta\theta}{d} - (\theta+\mu)\right)E.\end{aligned}$$

Since $\mathcal{R}_0 = \frac{\beta(\theta+\sigma_1 d)}{(\theta+\mu)d}$, we have

$$\beta\sigma_1 + \frac{\beta\theta}{d} = (\theta+\mu)\mathcal{R}_0.$$

Hence,

$$\dot{V} \leq (\theta + \mu)(\mathcal{R}_0 - 1)E.$$

Therefore, $\dot{V} \leq 0$ whenever $\mathcal{R}_0 < 1$, with equality if and only if $E = 0$. Since $\dot{I} = \theta E - dI$, it follows that $E = 0$ implies $I = 0$. Consequently, the largest positively invariant subset of $\{\dot{V} = 0\}$ is the DFE. According to LaSalle's invariance principle, every solution originating in the feasible region converges to the DFE. Hence, the DFE is globally asymptotically stable whenever $\mathcal{R}_0 < 1$. □

Theorem 6 (Uniform Persistence)**.** Assume that the feasible region Ω is positively invariant and that all solutions of the system are uniformly bounded. If $R_0 > 1$, the disease persists uniformly in the population. Consequently, there exists a positive constant $\Pi > 0$ such that every solution of the model with positive initial conditions satisfies

$$\liminf_{t\to\infty} S(t) \geq \Pi,$$

$$\liminf_{t\to\infty} E(t) \geq \Pi,$$

$$\liminf_{t\to\infty} I(t) \geq \Pi,$$

$$\liminf_{t\to\infty} H(t) \geq \Pi,$$

and

$$\liminf_{t\to\infty} R(t) \geq \Pi.$$

Proof. Consider the Lyapunov function

$$\begin{aligned} V =& q_1 \left(S - S^{**} - S^{**} \ln \frac{S}{S^{**}} \right) + q_2 \left(E - E^{**} - E^{**} \ln \frac{E}{E^{**}} \right) \\ &+ q_3 \left(I - I^{**} - I^{**} \ln \frac{I}{I^{**}} \right) + q_4 \left(H - H^{**} - H^{**} \ln \frac{H}{H^{**}} \right) \\ &+ q_5 \left(R - R^{**} - R^{**} \ln \frac{R}{R^{**}} \right), \end{aligned} \tag{10}$$

where $q_i > 0, \quad i = 1, 2, \ldots, 5$. Since

$$x - a - a\ln\left(\frac{x}{a}\right) \geq 0, \quad x > 0,\ a > 0,$$

the Lyapunov function satisfies $V(S, E, I, H, R) \geq 0$, with equality if and only if $(S, E, I, H, R) = (S^{**}, E^{**}, I^{**}, H^{**}, R^{**})$. This is precisely the endemic equilibrium.

Differentiating V along the solutions of the system gives

$$\begin{aligned}\frac{dV}{dt} =& q_1\left(1 - \frac{S^{**}}{S}\right)\frac{dS}{dt} + q_2\left(1 - \frac{E^{**}}{E}\right)\frac{dE}{dt} \\ &+ q_3\left(1 - \frac{I^{**}}{I}\right)\frac{dI}{dt} + q_4\left(1 - \frac{H^{**}}{H}\right)\frac{dH}{dt} \\ &+ q_5\left(1 - \frac{R^{**}}{R}\right)\frac{dR}{dt}.\end{aligned} \tag{11}$$

Substituting the governing equations into the above expression yields

$$\begin{aligned}\frac{dV}{dt} =& q_1\left(1 - \frac{S^{**}}{S}\right)\left[bN\left(1 - \frac{N}{K}\right) - \frac{\beta(I + \sigma_1 E)S}{N} - \mu S + \psi R\right] \\ &+ q_2\left(1 - \frac{E^{**}}{E}\right)\left[\frac{\beta(I + \sigma_1 E)S}{N} - (\theta + \mu)E\right] \\ &+ q_3\left(1 - \frac{I^{**}}{I}\right)\left[\theta E - (\delta_1 + \mu + \gamma_1 + \alpha)I\right] \\ &+ q_4\left(1 - \frac{H^{**}}{H}\right)\left[\alpha I - (\mu + \delta_2 + \gamma_2)H\right] \\ &+ q_5\left(1 - \frac{R^{**}}{R}\right)\left[\gamma_1 I + \gamma_2 H - (\mu + \psi)R\right].\end{aligned} \tag{12}$$

For convenience, we decompose the derivative into five components,

$$\frac{dV}{dt} = L_1 + L_2 + L_3 + L_4 + L_5,$$

where

$$L_1 = q_1\left(1-\frac{S^{**}}{S}\right)bN\left(1-\frac{N}{K}\right),$$

$$L_2 = -q_1\left(1-\frac{S^{**}}{S}\right)\frac{\beta(I+\sigma_1 E)S}{N} + q_2\left(1-\frac{E^{**}}{E}\right)\frac{\beta(I+\sigma_1 E)S}{N},$$

$$\begin{aligned} L_3 = &-q_1\mu\left(1-\frac{S^{**}}{S}\right)S - q_2(\theta+\mu)\left(1-\frac{E^{**}}{E}\right)E \\ &- q_3(\delta_1+\mu+\gamma_1+\alpha)\left(1-\frac{I^{**}}{I}\right)I - q_4(\mu+\delta_2+\gamma_2)\left(1-\frac{H^{**}}{H}\right)H \\ &- q_5(\mu+\psi)\left(1-\frac{R^{**}}{R}\right)R, \end{aligned}$$

$$L_4 = q_3\theta\left(1-\frac{I^{**}}{I}\right)E + q_4\alpha\left(1-\frac{H^{**}}{H}\right)I,$$

$$L_5 = q_1\psi\left(1-\frac{S^{**}}{S}\right)R + q_5\left(1-\frac{R^{**}}{R}\right)(\gamma_1 I+\gamma_2 H).$$

Hence,

$$\frac{dV}{dt} = L_1 + L_2 + L_3 + L_4 + L_5.$$

Because the feasible region Ω is positively invariant and bounded, every solution of the system remains positive and uniformly bounded. Consequently, there exists a positive constant M such that

$$0 < S(t),\, E(t),\, I(t),\, H(t),\, R(t),\, N(t) \le M, \quad t \ge 0. \tag{13}$$

Furthermore, all model parameters are positive constants, and the incidence function is continuously differentiable on Ω. Hence, every component of the vector field is continuous in the compact invariant region Ω.

Recall that $\frac{dV}{dt} = L_1 + L_2 + L_3 + L_4 + L_5$. Because each state variable remains bounded in the compact invariant region Ω, every function L_i, $i = 1, \ldots, 5$, is continuous on Ω. Therefore, by the Extreme Value Theorem, there exist positive constants C_1, C_2, C_3 and C_4 such that

$$L_1 \le C_1, \quad L_2 \le C_2, \quad L_4 \le C_3, \quad L_5 \le C_4.$$

Moreover,

$$\left(1-\frac{x^{**}}{x}\right)x = x - x^{**},$$

so that

$$\begin{aligned} L_3 = &- q_1\mu(S - S^{**}) - q_2(\theta + \mu)(E - E^{**}) \\ &- q_3(\delta_1 + \mu + \gamma_1 + \alpha)(I - I^{**}) \\ &- q_4(\mu + \delta_2 + \gamma_2)(H - H^{**}) \\ &- q_5(\mu + \psi)(R - R^{**}). \end{aligned}$$

Let

$$\begin{aligned} \eta = \min\Big\{ & q_1\mu,\ q_2(\theta + \mu),\ q_3(\delta_1 + \mu + \gamma_1 + \alpha), \\ & q_4(\mu + \delta_2 + \gamma_2),\ q_5(\mu + \psi) \Big\} > 0. \end{aligned}$$

Then,

$$L_3 \leq -\eta \left[(S - S^{**}) + (E - E^{**}) + (I - I^{**}) + (H - H^{**}) + (R - R^{**}) \right]. \tag{14}$$

Combining the above estimates yields

$$\frac{dV}{dt} \leq C - \eta \left[(S - S^{**}) + (E - E^{**}) + (I - I^{**}) + (H - H^{**}) + (R - R^{**}) \right], \tag{15}$$

where $C = C_1 + C_2 + C_3 + C_4$. Equation (15) shows that the derivative of the Lyapunov function is bounded above by a negative linear function of the deviation of the solution from the endemic equilibrium outside a compact neighborhood of the endemic equilibrium. □

5.2 Bifurcation analysis

The local stability result established that the DFE is locally asymptotically stable whenever $\mathcal{R}_0 < 1$. However, stability analysis alone does not describe the qualitative behavior of the model in a neighborhood of the critical threshold $\mathcal{R}_0 = 1$. In particular, it does not indicate whether the model undergoes a forward or backward bifurcation as the basic reproduction number passes through unity. To investigate the local dynamics near the critical threshold, the center manifold theory developed by Castillo-Chavez and Song [6] is employed. The transmission coefficient β was selected as the bifurcation parameter. The following theorem provides a criterion for determining the direction of the bifurcation.

Theorem 7 (Castillo-Chavez and Song)**.** Suppose that the Jacobian matrix of system (1) evaluated at the DFE has a simple zero eigenvalue when $\mathcal{R}_0 = 1$, and all the remaining eigenvalues have negative real parts. Let

$$a = \sum_{k,i,j=1}^{5} v_k w_i w_j \frac{\partial^2 f_k}{\partial x_i \partial x_j}(\mathcal{E}_0, \beta^*), \tag{16}$$

and

$$b = \sum_{k,i=1}^{5} v_k w_i \frac{\partial^2 f_k}{\partial x_i \partial \beta}(\mathcal{E}_0, \beta^*), \tag{17}$$

denote the center manifold coefficients, where $w = (w_1, \ldots, w_5)^T$ and $v = (v_1, \ldots, v_5)$ are the right and left eigenvectors associated with the simple zero eigenvalue, respectively. Then

1. if $a < 0$ and $b > 0$, the system undergoes a forward (supercritical transcritical) bifurcation;
2. if $a > 0$ and $b > 0$, the system undergoes a backward (subcritical) bifurcation.

Proof. The basic reproduction number obtained in section 4 is

$$\mathcal{R}_0 = \frac{\beta(\theta + \sigma_1 d)}{(\theta + \mu)d}, \tag{18}$$

where $d = \alpha + \delta_1 + \gamma_1 + \mu$. Substituting the bifurcation point $\mathcal{R}_0 = 1$ into (18) gives the critical transmission coefficient

$$\beta^* = \frac{(\theta + \mu)d}{\theta + \sigma_1 d}. \tag{19}$$

Hence, the DFE loses stability as the transmission coefficient passes through the critical value $\beta = \beta^*$. To facilitate the center manifold analysis, introduce the new variables

$$x_1 = S, \quad x_2 = E, \quad x_3 = I, \quad x_4 = H, \quad x_5 = R.$$

Then system (1) can be written in vector form as

$$\frac{dX}{dt} = F(X, \beta),$$

where $X = (x_1, x_2, x_3, x_4, x_5)^T$ and $F(X, \beta) = (f_1, f_2, f_3, f_4, f_5)^T$, with

$$f_1 = bN\left(1 - \frac{N}{K}\right) - \beta(x_3 + \sigma_1 x_2)\frac{x_1}{N} - \mu x_1 + \psi x_5,$$

$$f_2 = \beta(x_3 + \sigma_1 x_2)\frac{x_1}{N} - (\theta + \mu)x_2,$$

$$f_3 = \theta x_2 - d x_3,$$

$$f_4 = \alpha x_3 - k_3 x_4,$$
$$f_5 = \gamma_1 x_3 + \gamma_2 x_4 - k_4 x_5, \tag{20}$$

where $k_3 = \mu + \delta_2 + \gamma_2, \quad k_4 = \mu + \psi$. The DFE is

$$E_0 = \left(K\left(1 - \frac{\mu}{b}\right), 0, 0, 0, 0\right),$$

which satisfies $N = S^* = K\left(1 - \frac{\mu}{b}\right)$. Consequently, $\frac{S^*}{N^*} = 1$. The Jacobian matrix of system (20) evaluated at the DFE and at the critical transmission coefficient $\beta = \beta^*$ is

$$J(E_0, \beta^*) = \begin{pmatrix} \mu - b & -\beta^*\sigma_1 & -\beta^* & 0 & \psi \\ 0 & \beta^*\sigma_1 - k_1 & \beta^* & 0 & 0 \\ 0 & \theta & -d & 0 & 0 \\ 0 & 0 & \alpha & -k_3 & 0 \\ 0 & 0 & \gamma_1 & \gamma_2 & -k_4 \end{pmatrix}, \tag{21}$$

where $k_1 = \theta + \mu, \quad d = \alpha + \delta_1 + \gamma_1 + \mu, \quad k_3 = \mu + \delta_2 + \gamma_2, \quad k_4 = \mu + \psi$. The characteristic polynomial of the Jacobian matrix is obtained from

$$\chi(\lambda) = \det(\lambda I - J(\mathcal{E}_0, \beta^*)).$$

Since the Jacobian matrix possesses a block lower triangular structure, its determinant can be factorized as

$$\chi(\lambda) = (\lambda + b - \mu)(\lambda + k_3)(\lambda + k_4)\det(M), \tag{22}$$

where

$$M = \begin{pmatrix} \lambda - \beta^*\sigma_1 + k_1 & -\beta^* \\ -\theta & \lambda + d \end{pmatrix}.$$

Therefore,

$$\det(M) = (\lambda - \beta^*\sigma_1 + k_1)(\lambda + d) - \beta^*\theta. \tag{23}$$

Expanding (23) gives

$$\det(M) = \lambda^2 + (d + k_1 - \beta^*\sigma_1)\lambda + d(k_1 - \beta^*\sigma_1) - \beta^*\theta.$$

Consequently,

$$\begin{aligned}\chi(\lambda) &= (\lambda + b - \mu)(\lambda + k_3)(\lambda + k_4) \\ &\times \left[\lambda^2 + (d + k_1 - \beta^*\sigma_1)\lambda + d(k_1 - \beta^*\sigma_1) - \beta^*\theta\right].\end{aligned} \tag{24}$$

To verify the existence of a simple zero eigenvalue, let $\lambda = 0$, then $\det(M) = d(k_1 - \beta^*\sigma_1) - \beta^*\theta$. Using $\beta^* = \frac{k_1 d}{\theta + \sigma_1 d}$, we obtain

$$\begin{aligned} d(k_1 - \beta^*\sigma_1) - \beta^*\theta &= dk_1 - \frac{k_1 d^2 \sigma_1}{\theta + \sigma_1 d} - \frac{k_1 d\theta}{\theta + \sigma_1 d} \\ &= dk_1 - \frac{k_1 d(\sigma_1 d + \theta)}{\theta + \sigma_1 d} \\ &= dk_1 - dk_1 \\ &= 0. \end{aligned}$$

Hence, $\lambda = 0$ is the eigenvalue of the Jacobian matrix. Next, $\mu - b < 0$, since the DFE exists whenever $b > \mu$. Similarly, $-k_3 < 0$, $-k_4 < 0$, because every parameter is positive. The remaining eigenvalue associated with the quadratic factor satisfies

$$\lambda = -(d + k_1 - \beta^*\sigma_1),$$

which is negative because $d + k_1 - \beta^*\sigma_1 = d + \theta + \mu - \beta^*\sigma_1 > 0$. Therefore, the Jacobian matrix possesses exactly one simple zero eigenvalue, whereas all the remaining eigenvalues have strictly negative real parts. Hence, all hypotheses of the Castillo–Chavez and Song theorem are satisfied. □

Computation of the right eigenvector

Let $w = (w_1, w_2, w_3, w_4, w_5)^T$ be the right eigenvector associated with the simple zero eigenvalue of Jacobian matrix, then $J(\mathcal{E}_0, \beta^*)w = 0$. From (21), we obtain the following system of linear equations

$$(\mu - b)w_1 - \beta^*\sigma_1 w_2 - \beta^* w_3 + \psi w_5 = 0, \tag{25}$$

$$(\beta^*\sigma_1 - k_1)w_2 + \beta^* w_3 = 0, \tag{26}$$

$$\theta w_2 - d w_3 = 0, \tag{27}$$

$$\alpha w_3 - k_3 w_4 = 0, \tag{28}$$

$$\gamma_1 w_3 + \gamma_2 w_4 - k_4 w_5 = 0. \tag{29}$$

Equation (27) immediately gives

$$w_3 = \frac{\theta}{d} w_2. \tag{30}$$

Substituting (30) into (28) yields

$$w_4 = \frac{\alpha\theta}{dk_3} w_2. \tag{31}$$

Similarly, substituting (30) and (31) into (29) gives

$$w_5 = \frac{\theta}{dk_4}\left(\gamma_1 + \frac{\alpha\gamma_2}{k_3}\right) w_2. \tag{32}$$

Finally, substituting (30) and (32) into (25) gives

$$(\mu - b)w_1 = \beta^*\sigma_1 w_2 + \frac{\beta^*\theta}{d} w_2 - \frac{\psi\theta}{dk_4}\left(\gamma_1 + \frac{\alpha\gamma_2}{k_3}\right) w_2.$$

Hence,

$$w_1 = \frac{\beta^*\left(\sigma_1 + \frac{\theta}{d}\right) - \frac{\psi\theta}{dk_4}\left(\gamma_1 + \frac{\alpha\gamma_2}{k_3}\right)}{\mu - b} w_2. \tag{33}$$

Therefore, the right eigenvector corresponding to the simple zero eigenvalue is

$$w = \begin{pmatrix} \dfrac{\beta^*\left(\sigma_1 + \frac{\theta}{d}\right) - \frac{\psi\theta}{dk_4}\left(\gamma_1 + \frac{\alpha\gamma_2}{k_3}\right)}{\mu - b} w_2 \\ w_2 \\ \frac{\theta}{d} w_2 \\ \frac{\alpha\theta}{dk_3} w_2 \\ \frac{\theta}{dk_4}\left(\gamma_1 + \frac{\alpha\gamma_2}{k_3}\right) w_2 \end{pmatrix},$$

where w_2 is arbitrary.

Computation of the left eigenvector

Let $v = (v_1, v_2, v_3, v_4, v_5)$ be the left eigenvector associated with the simple zero eigenvalue, then $vJ(\mathcal{E}_0, \beta^*) = 0$, or equivalently, $J(\mathcal{E}_0, \beta^*)^T v = 0$. This gives

$$(\mu - b)v_1 = 0, \tag{34}$$

$$-\beta^*\sigma_1 v_1 + (\beta^*\sigma_1 - k_1)v_2 + \theta v_3 = 0, \tag{35}$$

$$-\beta^* v_1 + \beta^* v_2 - dv_3 + \alpha v_4 + \gamma_1 v_5 = 0, \tag{36}$$

$$-k_3 v_4 + \gamma_2 v_5 = 0, \tag{37}$$

$$\psi v_1 - k_4 v_5 = 0. \tag{38}$$

Since $\mu - b \neq 0$, (34) gives $v_1 = 0$. Substituting into (38) yields $v_5 = 0$. Equation (37) therefore becomes $v_4 = 0$.

Substituting these results into (36) gives $\beta^* v_2 - dv_3 = 0$, from which

$$v_3 = \frac{\beta^*}{d} v_2. \tag{39}$$

Finally, substituting (39) into (35) gives

$$\left(\beta^*\sigma_1 - k_1 + \frac{\beta^*\theta}{d}\right) v_2 = 0.$$

Since $\beta^* = \frac{k_1 d}{\theta + \sigma_1 d}$, we have

$$\beta^*\sigma_1 - k_1 + \frac{\beta^*\theta}{d} = 0.$$

Hence, Equation (35) is satisfied identically, implying that v_2 is arbitrary. Therefore, the left eigenvector associated with the simple zero eigenvalue is

$$v = \left(0,\, v_2,\, \frac{\beta^*}{d} v_2,\, 0,\, 0\right),$$

where v_2 is arbitrary.

Second-order partial derivatives

Because the functions f_3, f_4 and f_5 are linear, all their second-order partial derivatives vanish identically. Consequently, only the nonlinear functions f_1 and f_2 contribute to the computation of the center-manifold coefficients.

The nonlinear incidence term is

$$\Phi(S,E,I,H,R)=\beta\frac{S(I+\sigma_1 E)}{N},$$

where $N=S+E+I+H+R$. The required second-order partial derivatives are evaluated at the DFE

$$\mathcal{E}_0=(S^*,0,0,0,0),$$

where $S^*=N^*$. The second-order partial derivatives of the vector field are required to compute the center manifold coefficients. Because the functions f_3, f_4 and f_5 are linear, all their second-order partial derivatives vanish identically. Therefore, only the nonlinear functions f_1 and f_2 contribute to the bifurcation analysis. The nonlinear incidence term is $\Phi(x)=\beta\frac{x_1(x_3+\sigma_1 x_2)}{N}$, where $N=x_1+x_2+x_3+x_4+x_5$. The required derivatives are evaluated at the DFE

$$\mathcal{E}_0=(S^*,0,0,0,0),\quad S^*=N^*=K\left(1-\frac{\mu}{b}\right).$$

Since $f_2=\Phi-k_1x_2$, The linear term contributes no second-order derivatives. Differentiating the incidence function and evaluating at the DFE yields

$$\frac{\partial^2 f_2}{\partial x_1\partial x_2}(\mathcal{E}_0)=0, \tag{40}$$
$$\frac{\partial^2 f_2}{\partial x_1\partial x_3}(\mathcal{E}_0)=0, \tag{41}$$
$$\frac{\partial^2 f_2}{\partial x_2^2}(\mathcal{E}_0)=-\frac{2\beta^*\sigma_1}{S^*}, \tag{42}$$
$$\frac{\partial^2 f_2}{\partial x_2\partial x_3}(\mathcal{E}_0)=-\frac{\beta^*(1+\sigma_1)}{S^*}, \tag{43}$$
$$\frac{\partial^2 f_2}{\partial x_3^2}(\mathcal{E}_0)=-\frac{2\beta^*}{S^*}. \tag{44}$$

Since $f_1=bN\left(1-\frac{N}{K}\right)-\Phi-\mu x_1+\psi x_5$, The nonlinear infection term in f_1 is the negative of that appearing in f_2. Consequently,

$$\frac{\partial^2 f_1}{\partial x_1\partial x_2}(\mathcal{E}_0)=0, \tag{45}$$
$$\frac{\partial^2 f_1}{\partial x_1\partial x_3}(\mathcal{E}_0)=0, \tag{46}$$
$$\frac{\partial^2 f_1}{\partial x_2^2}(\mathcal{E}_0)=\frac{2\beta^*\sigma_1}{S^*}, \tag{47}$$

$$\frac{\partial^2 f_1}{\partial x_2 \partial x_3}(\mathcal{E}_0) = \frac{\beta^*(1+\sigma_1)}{S^*}, \tag{48}$$

$$\frac{\partial^2 f_1}{\partial x_3^2}(\mathcal{E}_0) = \frac{2\beta^*}{S^*}. \tag{49}$$

Moreover,

$$\frac{\partial^2 f_k}{\partial x_i \partial x_j}(\mathcal{E}_0) = 0, \quad k = 3, 4, 5,$$

for all $i, j = 1, \ldots, 5$. Hence, the nonzero second-order partial derivatives required in the computation of the Castillo–Chavez coefficients are given by (42)–(44) and (47)–(49).

Computation of the coefficient a

Following Castillo–Chavez and Song [6], the coefficient a is given by

$$a = \sum_{k,i,j=1}^{5} v_k w_i w_j \frac{\partial^2 f_k}{\partial x_i \partial x_j}(\mathcal{E}_0, \beta^*). \tag{50}$$

Since $v_1 = v_4 = v_5 = 0$, and the functions f_3, f_4 and f_5 are linear, their second-order partial derivatives vanish identically. Hence, only the nonlinear infection equation f_2 contributes to the coefficient a. Therefore,

$$a = v_2 \sum_{i,j=1}^{5} w_i w_j \frac{\partial^2 f_2}{\partial x_i \partial x_j}(\mathcal{E}_0, \beta^*). \tag{51}$$

From the Hessian matrix evaluated at the DFE, the only nonzero second-order partial derivatives are

$$\frac{\partial^2 f_2}{\partial x_2^2} = -\frac{2\beta^*\sigma_1}{S^*},$$

$$\frac{\partial^2 f_2}{\partial x_2 \partial x_3} = \frac{\partial^2 f_2}{\partial x_3 \partial x_2} = -\frac{\beta^*(1+\sigma_1)}{S^*},$$

and

$$\frac{\partial^2 f_2}{\partial x_3^2} = -\frac{2\beta^*}{S^*}.$$

Consequently,

$$a = v_2 \left[w_2^2 \frac{\partial^2 f_2}{\partial x_2^2} + 2 w_2 w_3 \frac{\partial^2 f_2}{\partial x_2 \partial x_3} + w_3^2 \frac{\partial^2 f_2}{\partial x_3^2} \right]. \tag{52}$$

Since $w_3 = \frac{\theta}{d} w_2$, it follows that

$$a = v_2 w_2^2 \left[\frac{\partial^2 f_2}{\partial x_2^2} + 2\frac{\theta}{d}\frac{\partial^2 f_2}{\partial x_2 \partial x_3} + \frac{\theta^2}{d^2}\frac{\partial^2 f_2}{\partial x_3^2} \right]. \tag{53}$$

Substituting the corresponding second-order partial derivatives gives

$$\begin{aligned} a &= v_2 w_2^2 \left[-\frac{2\beta^* \sigma_1}{S^*} - 2\frac{\theta}{d}\left(\frac{\beta^*(1+\sigma_1)}{S^*} \right) - \frac{\theta^2}{d^2}\left(\frac{2\beta^*}{S^*} \right) \right] \\ &= -\frac{2\beta^* v_2 w_2^2}{S^*} \left[\sigma_1 + \frac{\theta(1+\sigma_1)}{d} + \frac{\theta^2}{d^2} \right]. \end{aligned} \tag{54}$$

Since all model parameters are positive,

$$\beta^* > 0, \quad S^* > 0, \quad d > 0, \quad v_2 > 0, \quad w_2 > 0,$$

and

$$\sigma_1 + \frac{\theta(1+\sigma_1)}{d} + \frac{\theta^2}{d^2} > 0,$$

it follows immediately that $a < 0$. The coefficient b is defined by

$$b = \sum_{k,i=1}^{5} v_k w_i \frac{\partial^2 f_k}{\partial x_i \partial \beta}(\mathcal{E}_0, \beta^*). \tag{55}$$

Since $v_1 = v_4 = v_5 = 0$ and the functions f_3, f_4, and f_5 do not depend explicitly on the bifurcation parameter β, only the infected equation f_2 contributes to the above equation. Hence,

$$b = v_2 \sum_{i=1}^{5} w_i \frac{\partial^2 f_2}{\partial x_i \partial \beta}(\mathcal{E}_0, \beta^*). \tag{56}$$

The infection equation is

$$f_2 = \beta \frac{x_1(x_3 + \sigma_1 x_2)}{N} - k_1 x_2.$$

Differentiating with respect to β gives

$$\frac{\partial f_2}{\partial \beta} = \frac{x_1(x_3 + \sigma_1 x_2)}{N}.$$

Differentiating with respect to the state variables and evaluating at the DFE yields

$$\frac{\partial^2 f_2}{\partial x_1 \partial \beta}(\mathcal{E}_0, \beta^*) = 0,$$

$$\frac{\partial^2 f_2}{\partial x_2 \partial \beta}(\mathcal{E}_0, \beta^*) = \sigma_1,$$

$$\frac{\partial^2 f_2}{\partial x_3 \partial \beta}(\mathcal{E}_0, \beta^*) = 1,$$

while

$$\frac{\partial^2 f_2}{\partial x_4 \partial \beta} = \frac{\partial^2 f_2}{\partial x_5 \partial \beta} = 0.$$

Therefore,

$$b = v_2 \left(\sigma_1 w_2 + w_3\right). \tag{57}$$

Using the right eigenvector relation $w_3 = \frac{\theta}{d} w_2$, we obtain

$$b = v_2 w_2 \left(\sigma_1 + \frac{\theta}{d}\right). \tag{58}$$

Since

$$v_2 > 0, \qquad w_2 > 0, \qquad \sigma_1 > 0, \qquad \theta > 0, \qquad d > 0,$$

it follows immediately that $b > 0$. Therefore, $a < 0, \quad b > 0$. Hence, by the Castillo–Chavez and Song theorem, the proposed epidemic model undergoes a *forward (supercritical transcritical) bifurcation* at the critical threshold $\mathcal{R}_0 = 1$. Consequently, the DFE is locally asymptotically stable whenever $\mathcal{R}_0 < 1$, whereas a unique locally asymptotically stable endemic equilibrium emerges for $\mathcal{R}_0 > 1$.

6 Sensitivity analysis

Sensitivity analysis plays a vital role in determining which parameters should be prioritized when designing intervention strategies to reduce and control the spread of meningococcal meningitis. This analysis measures how variations in the basic reproduction number $(\mathcal{R}_0)$ correspond to proportional changes in individual parameters. In this study, the influence of each model parameter on $(\mathcal{R}_0)$ was examined at its given value. The analytical expression used to evaluate the sensitivity of $(\mathcal{R}_0)$ with respect to any parameter (τ) is given as:

$$\Upsilon_{\tau}^{\mathcal{R}_0} = \frac{\tau}{\mathcal{R}_0} \times \frac{\partial \mathcal{R}_0}{\partial \tau}.$$

Here, (τ) represents any arbitrary parameter that influences $(\mathcal{R}_0)$. By substituting the baseline parameter values listed in Table 2, the numerical sensitivity indices of model (1) were computed using $(\mathcal{R}_0)$ as the response variable. The outcomes of the sensitivity analysis are shown in Figure 2 and Table 3.

Table 3: Sensitivity indices of the model parameters

Parameter	Sensitivity Index
β	1
θ	-0.45635
μ	-0.000245421
σ_1	0.456545
α	-0.388146
δ_1	-0.0258764
γ_1	-0.129382

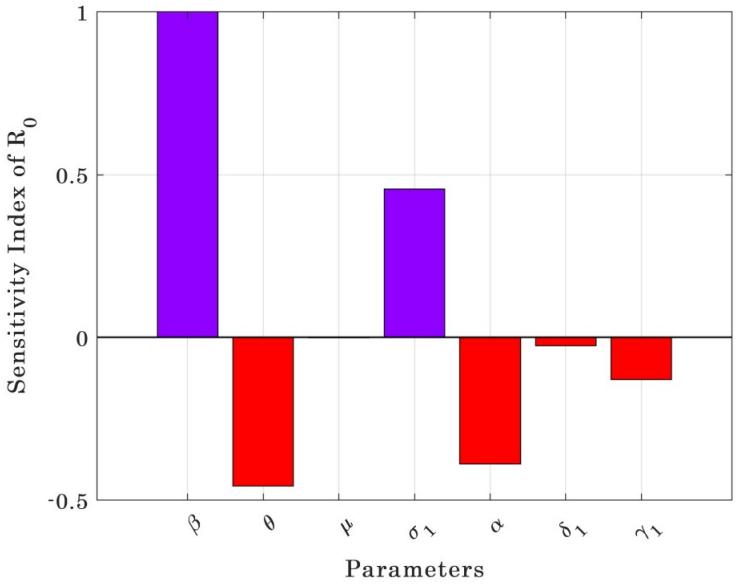
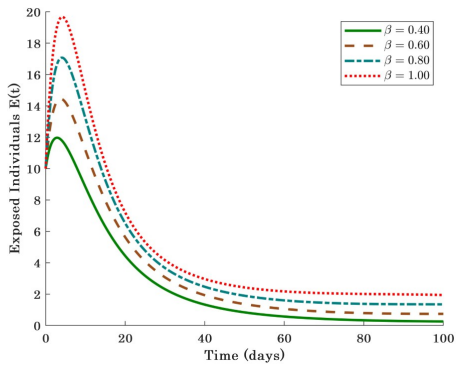


Figure 2: Sensitivity index of the basic reproduction number $\mathcal{R}_0$

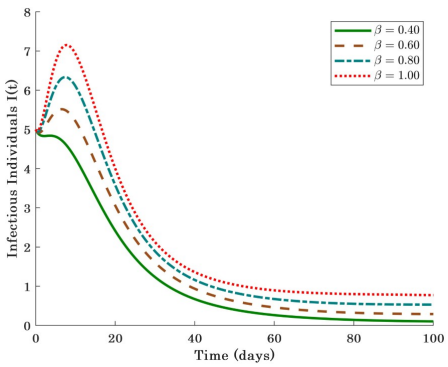
7 Numerical simulation

In this section, the proposed model (1) is numerically solved using the `ode45` solver in MATLAB®. The initial conditions are assumed as $S(0) = 1000$, $E(0) = 10$, $I(0) = 5$, $H(0) = 2$, $R(0) = 1$, $D(0) = 0$.

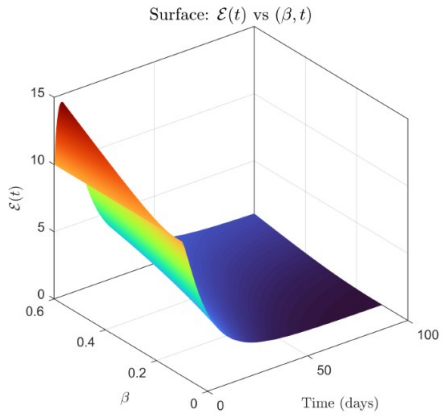
Figure 3 illustrates the impact of varying transmission rates (β) on the dynamics of exposed and infected human populations. The plot of the exposed group $E(t)$ for $\beta = 0.40, 0.60, 0.80,$ and 1.00 is shown in Figure 3(a). It can be observed that as the transmission rate increases, the number of exposed individuals rises sharply, reaching a higher peak before gradually declining over time. This behavior indicates that higher transmission rates accelerate the spread of

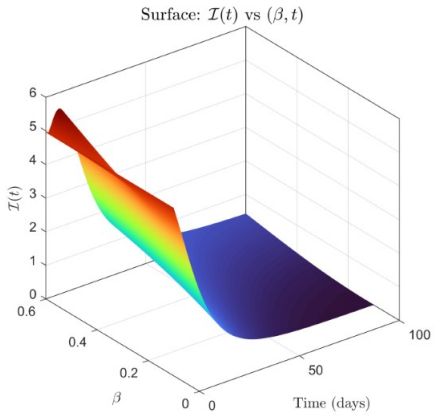


(a) Exposed Individuals $E(t)$



(b) Infected Individuals $I(t)$

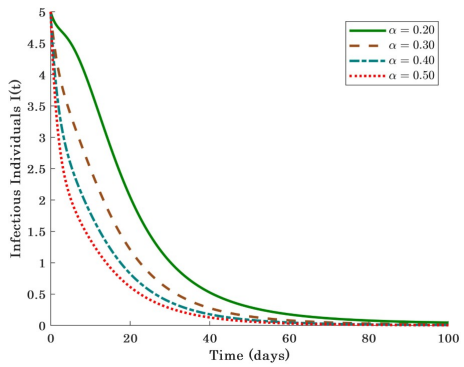




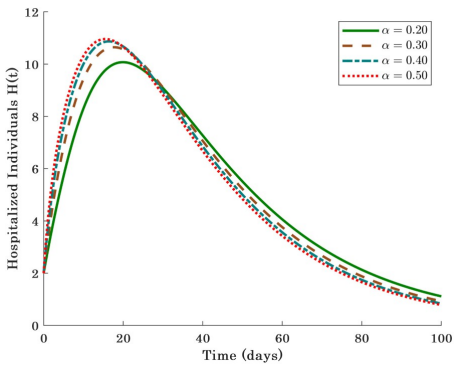
(c) Surface representation of the effects of β on $E(t)$ and $I(t)$

Figure 3: Impact of the transmission rate β on exposed and infected individuals

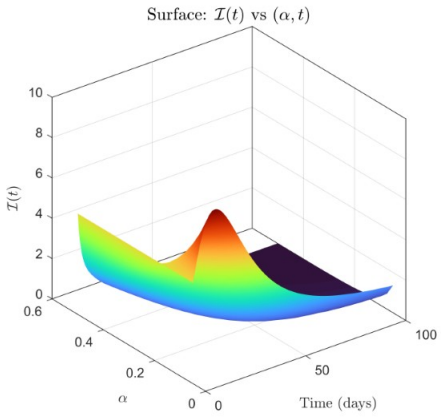
meningococcal meningitis, leading to increased exposure during the early stages of an outbreak. However, with effective hospitalization and treatment measures, the exposed population eventually decreases, contributing to the stabilization of the epidemic. Similarly, Figure 3(b) shows that the number of infected individuals $I(t)$ increases as β increases. The trajectories reveal that higher values of β lead to earlier and larger infection peaks, followed by a steady decline. This trend implies that elevated transmission rates hasten the progression from exposure to infection and prolong the active infection period until recovery or intervention reduces the case count. Figure 3(c) shows a surface plot summarizing these scenarios. Overall, the results depicted in Figure 3 emphasize that the transmission rate β plays a crucial role in shaping both the intensity and duration of meningococcal meningitis outbreaks. While higher β values intensify and

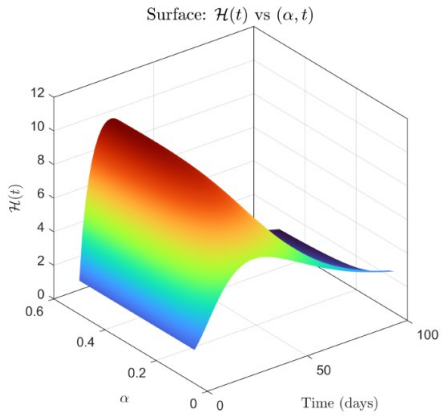


(a) Infected Individuals $I(t)$



(b) Hospitalized Individuals $H(t)$



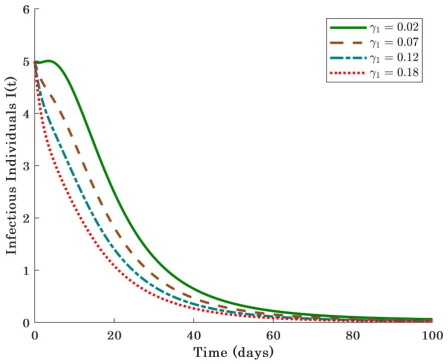


(c) Surface representation of the effects of α on $I(t)$ and $H(t)$

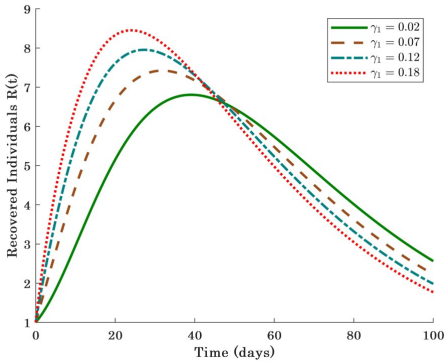
Figure 4: Impact of the hospitalization rate α on infected and hospitalized individuals

accelerate disease spread, timely medical responses and treatment interventions help mitigate infections and restore system equilibrium over time.

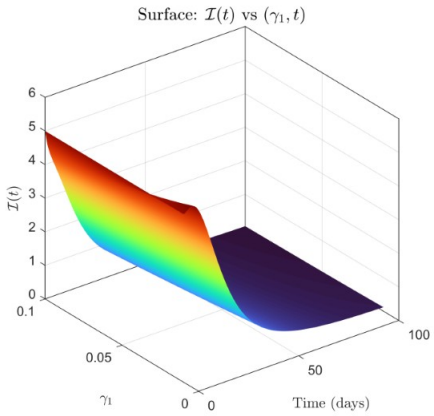
Figure 4 illustrates the influence of varying hospitalization rates (α), which represent the cure rate of infected individuals ($\alpha > 0$), on the dynamics of the infected and hospitalized human populations. Simulations were conducted for $\alpha = 0.20, 0.30, 0.40$, and 0.50. Figure 4(a) depicts the temporal evolution of the infected population $I(t)$. It is observed that as the hospitalization rate α increases, the number of infected individuals decreases. A higher value of α implies faster hospitalization and recovery, thereby reducing the proportion of actively infected individuals within the population. Consequently, the infection curves corresponding to larger α values decrease more sharply, indicating a faster containment of the disease. The temporal

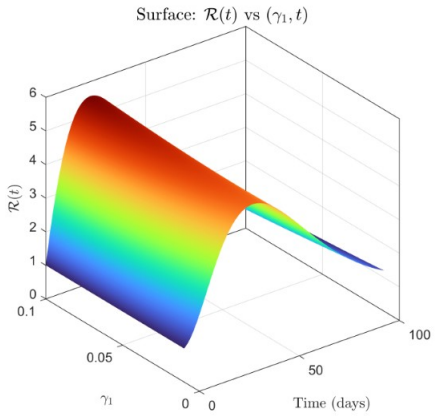


(a) Infected Individuals $I(t)$



(b) Recovered Individuals $R(t)$



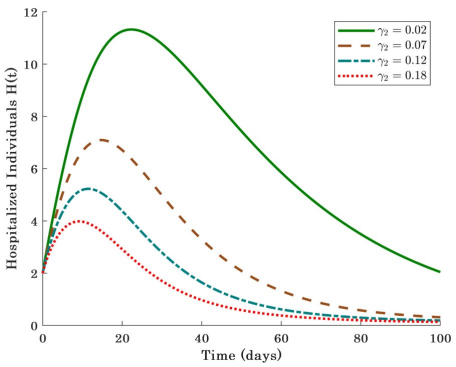


(c) Surface representation of the effects of γ_1 on $I(t)$ and $R(t)$

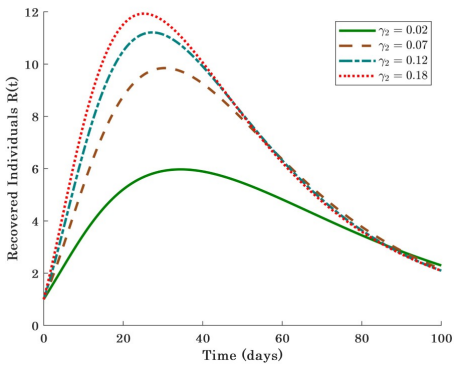
Figure 5: Impact of the recovery rate of infectious individuals γ_1 on infected and recovered individuals

behavior of hospitalized individuals $H(t)$ is shown in Figure 4(b). Initially, a higher α resulted in an earlier and slightly elevated peak in hospitalizations due to the faster transfer of infected individuals to treatment facilities. Over time, however, the hospitalized population diminishes as recovery rates improve with increasing α. Figure 4(c) shows a surface representation of these outcomes. Overall, the results in Figure 4 demonstrate that the hospitalization rate α plays a vital role in controlling the spread of meningococcal meningitis cases. A higher α accelerates the hospitalization and recovery of infected individuals and contributes to stabilizing the epidemic and reducing the overall disease prevalence in the long term.

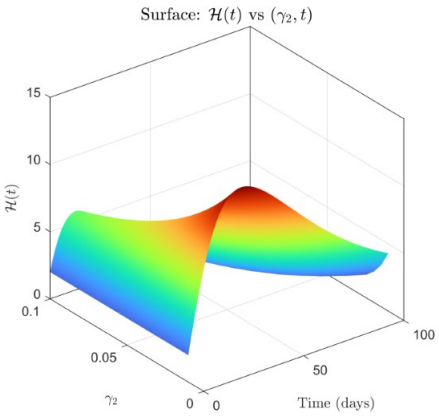
Figure 5 demonstrates how variations in the recovery rate (γ_1) of infectious individuals influence the dynamics of both the infected and recovered populations. The simulations were

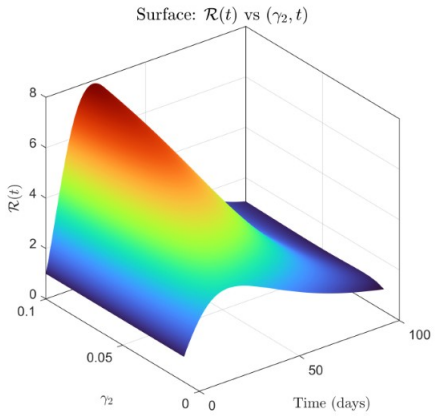


(a) Hospitalized Individuals $H(t)$



(b) Recovered Individuals $R(t)$

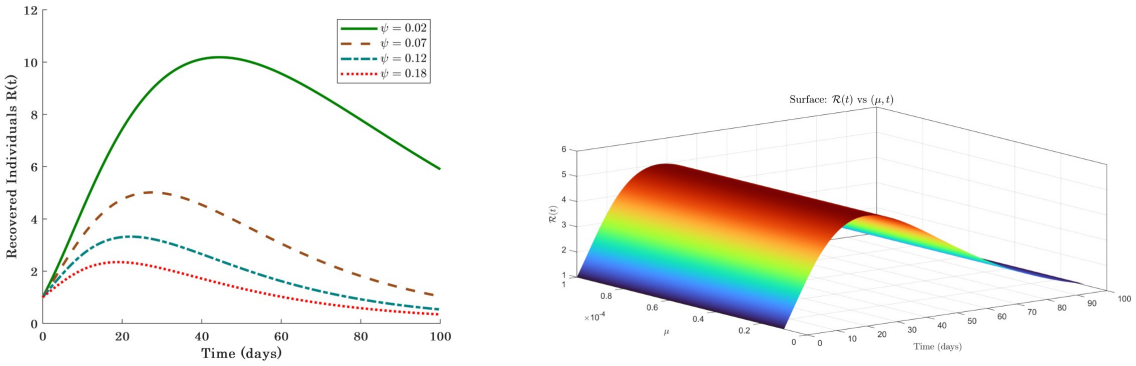




(c) Surface representation of the effects of γ_2 on $H(t)$ and $R(t)$

Figure 6: Impact of the recovery rate of hospitalized individuals γ_2 on hospitalized and recovered individuals

performed for $\gamma_1 = 0.02, 0.07, 0.12$, and 0.18. Figure 5(a) shows the temporal evolution of the number of infected individuals $I(t)$ for various recovery rates. It is evident that as γ_1 increases, the number of infected individuals declines more rapidly. Lower γ_1 values corresponded to slower recovery, resulting in a prolonged infection period and a delayed reduction in cases. In contrast, higher recovery rates led to a faster decrease in infections, indicating that individuals recovered more quickly and the disease was brought under control sooner. These results highlight that enhancing the recovery rate significantly accelerates infection suppression and shortens the infectious period of the disease. Figure 5(b) illustrates the corresponding changes in the recovered population $R(t)$ for different values of γ_1. As the recovery rate increased, the recovered population grew more steeply, reaching a higher peak in a shorter period. This pattern implies that a larger



(a) Recovered individuals $R(t)$ (b) Surface representation of the effects of μ on $R(t)$

Figure 7: Impact of the loss of immunity rate ψ on recovered individuals

fraction of infected individuals transitions into the recovered class earlier when recovery is more efficient. After peaking, the number of recovered individuals gradually stabilizes, signifying that the outbreak is under control as recovery surpasses new infection. Figure 5(c) presents a surface plot that visually summarizes these dynamics. Overall, the results in Figure 5 underscore that the recovery rate γ_1 is a crucial determinant of the progression and mitigation of meningococcal meningitis. A higher γ_1 not only reduces the number of active infections but also enhances the accumulation of recovered individuals, thereby promoting faster epidemic stabilization and improving the disease management.

Figure 6 depicts the influence of the recovery rate of hospitalized individuals (γ_2) on the variation in both hospitalized ($H(t)$) and recovered ($R(t)$) populations. Simulations were performed for different recovery rate values: $\gamma_2 = 0.02, 0.07, 0.12$, and 0.18. Figure 6(a) illustrates the temporal behavior of hospitalized individuals $H(t)$ for the given recovery rates. It is observed that as γ_2 increases, the number of hospitalized individuals decreases noticeably. At lower recovery rates (e.g., $\gamma_2 = 0.02$), hospitalization levels reach a higher peak and decline slowly over time, reflecting prolonged hospital stays and delayed recovery. Conversely, at higher recovery rates (e.g., $\gamma_2 = 0.18$), hospitalization peaks are smaller and decline more rapidly, indicating a faster recovery and shorter hospitalization durations. These observations highlight that an increased recovery rate effectively reduces the overall hospitalization burden by facilitating quicker patient discharge. The dynamics of the recovered population $R(t)$ under the same recovery rate variation are shown in Figure 6(b). As γ_2 increases, the number of recovered individuals increases more sharply, attaining a higher peak in a shorter time frame. This indicates that higher recovery rates among hospitalized patients directly accelerate the accumulation of recoveries within the population. Following the peak, the recovered population stabilizes gradually, signifying that the

epidemic subsides as recovery becomes more dominant than new admissions. Figure 6(c) provides a surface representation that visually consolidates these observations. Overall, the results in Figure 6 emphasize the crucial role of the recovery rate γ_2 in shaping both hospitalization duration and recovery trends. A higher γ_2 not only shortens hospital stays and promotes faster patient recovery but also alleviates pressure on healthcare systems and contributes significantly to the effective control of meningococcal meningitis.

Figure 7 demonstrates the influence of the loss of immunity rate (ψ) on the dynamics of the recovered population ($R(t)$) in a human community. Simulations were conducted for different values of $\psi = 0.02, 0.07, 0.12$, and 0.18. As shown in Figure 7(a), the number of recovered individuals decreases notably as the loss of immunity rate ψ increases. For lower ψ values (e.g., $\psi = 0.02$), the recovered population reaches a higher peak and remains elevated for a longer period, indicating prolonged immunity and fewer individuals reverting to the susceptible class. Conversely, at higher ψ values (e.g., $\psi = 0.18$), the recovered curve displays a smaller peak followed by a faster decline, signifying a quicker waning of immunity and a faster return of recovered individuals to being susceptible. These observations suggest that an elevated loss of immunity reduces the duration of post-recovery protection, thereby accelerating the transition of recovered individuals back into the susceptible group. This process diminishes the long-term resistance of the population to infections and increases the likelihood of disease resurgence. Figure 7(b) presents a surface plot that visually summarizes these relationships. In summary, the results in Figure 7 highlight that the loss of immunity rate ψ plays a critical role in determining the persistence and stability of recovery. Lower ψ values support sustained immunity and a larger recovered population over time, whereas higher ψ values weaken recovery durability, potentially leading to repeated meningococcal meningitis outbreaks.

8 Conclusion

This study investigated the transmission dynamics of meningococcal meningitis using a novel compartmental model that incorporates hospitalization and treatment effects. The theoretical behavior of the model was examined using the threshold dynamics. The model incorporates key epidemiological parameters, such as the intrinsic per capita recruitment rate (b), carrying capacity (K), natural death rate (μ), and transmission rate (β), which characterize the demographic and epidemiological features of the disease. The progression rate (θ) governs the transition from the exposed class to the infectious class, whereas the modification parameter (σ_1) accounts for the reduced infectiousness of exposed individuals. Disease progression and clinical outcomes are

further described by the hospitalization rate (α), recovery rates of infectious and hospitalized individuals (γ_1 and γ_2), and disease-induced death rates (δ_1 and δ_2).

Numerical simulations demonstrated that increasing the hospitalization and recovery rates substantially reduced the number of infectious individuals and shortened the duration of the epidemic. Furthermore, the sensitivity analysis identified the transmission, hospitalization, and recovery rates as the most influential parameters affecting the basic reproduction number ($\mathcal{R}_0$). These findings highlight the important role of timely hospitalization and effective clinical management in reducing disease transmission. Overall, the proposed framework provides valuable insights into the transmission dynamics of meningococcal meningitis and offers a useful tool for evaluating disease-control strategies. Future work may extend the present model by incorporating vaccination, delay differential equations to capture time-dependent epidemiological processes, stochastic effects, and optimal control strategies to design more effective intervention policies.

Author contributions

Idowu Oluwarimi Joshua: Conceptualization, Writing-original draft, Investigation, Validation, and Writing-review and Editing.
K. Kiruthika: Writing-original draft, Investigation, Validation, Writing-review and Editing.
Mehmet Gümüş: Investigation, Validation, Writing-review and Editing.
Ali Raza: Conceptualization, Writing-original draft, Investigation, Validation, and Writing-review and Editing.
E. Chandrakumar: Investigation, Validation, Writing-review and Editing.
N. Jeeva: Conceptualization, Writing-original draft, Software, Investigation, Validation, Supervision and Writing-review and Editing.

Funding

This research did not receive any external funding.

Data Availability Statement

All data used are available within the manuscript.

Ethics Statement

This study did not require ethical approval.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Afolabi, M.A., Adewoye, K.S., Folorunso, A.I. and Omoloye, M.A. *A mathematical model on transmission dynamics of meningococcal meningitis*, Iconic Res. Eng. J. (2021), 59–66.

[2] Agusto, F. and Leite, M. *Optimal control and cost-effective analysis of the 2017 meningitis outbreak in Nigeria*, Infect. Dis. Model., 4 (2019), 161–187.

[3] Anderson, R.M. and May, R.M. *Infectious Diseases of Humans: Dynamics and Control*, Oxford Univ. Press, 1991.

[4] Asamoah, J.K.K., Nyabadza, F., Seidu, B., Chand, M. and Dutta, H. *Mathematical modelling of bacterial meningitis transmission dynamics with control measures*, Comput. Math. Methods Med. (2018), 1–21.

[5] Bani-Yaghoub, M., Gautam, R., Shuai, Z., van den Driessche, P. and Ivanek, R. *Reproduction numbers for infections with free-living pathogens growing in the environment*, J. Biol. Dyn., 6 (2012), 923–940.

[6] Castillo-Chavez, C. and Song, B. *Dynamical models of tuberculosis and their applications*, Math. Biosci. Eng., 1(2) (2004), 361–404.

[7] Chen, H., Li, M., Tu, S., Zhang, X., Wang, X., Zhang, Y., Zhao, C., Guo, Y. and Wang, H. *Metagenomic data from cerebrospinal fluid permits tracing the origin and spread of Neisseria meningitidis CC4821 in China*, Commun. Biol., 5 (2022), 839.

[8] Chen, M., Zhang, C., Zhang, X. and Chen, M. *Meningococcal quinolone resistance originated from several commensal Neisseria species*, Antimicrob. Agents Chemother., 64 (2020) 10–128.

[9] Coen, P.G., Cartwright, K. and Stuart, J. *Mathematical modelling of infection and disease due to Neisseria meningitidis and Neisseria lactamica*, Int. J. Epidemiol., 29 (2000), 180–188.

[10] Diekmann, O. and Hesterbeek, J. *Mathematical Epidemiology of Infectious Diseases*, Wiley, 2000.

[11] Fresnadillo Martínez, M.J., García Merino, E., García Sánchez, E., García Sánchez, J.E., Martín del Rey, A. and Rodríguez Sánchez, G. *Mathematical model to study the meningococcal meningitis*, Procedia Comput. Sci., 18 (2013), 2492–2495.

[12] Lakshmi, G.K., Dharmalingam, K.M. and Jeeva, N. *Numerical analysis of dengue virus transmission under integer-order and fractional-order frameworks*, Netw. Model. Anal. Health Inform. Bioinform., 15(1) (2026) 145.

[13] Ibrahim, H., Maksod, S.A., Khorshed, M., Alsisi, A., Fouad, N., Rady, H., Al Amir, A., Mohamed, A. and Hosny, M. *Meningococcal meningitis, a life-threatening disease with a dangerous skin rash: Two case reports*, Egypt. J. Intern. Med., 35 (2023), 63.

[14] Isa, R.S., Mensah, K., Acheneje, G.O., Jeeva, N., Gümüş, M., Ali, K. and Raza, A. *Modeling corruption as a social contagion: A mathematical approach to dynamics, thresholds, and control*, Int. J. Dyn. Control, 14(7) (2026) 259.

[15] Jeeva, N. and Manivel, M. *Fractional dynamics of type 2 diabetes: A memory-dependent compartmental model for prevention and control*, Franklin Open, 16 (2026), 100696.

[16] Johri, S., Gorthi, S.P. and Anand, A.C. *Meningococcal meningitis*, Med. J. Armed Forces India, 61(4) (2005), 369–374.

[17] Karachaliou, A., Conlan, A.J.K., Preziosi, M.-P. and Trotter, C.L. *Modeling long-term immunity against meningococcal disease: Implications for vaccine strategies*, Epidemics, 11 (2015), 1–9.

[18] Keeling, M.J. and Rohani, P. *Modeling Infectious Diseases in Humans and Animals*, Princeton Univ. Press, Princeton, NJ, 2008.

[19] Kotola, B.S. and Mekonnen, T.T. *Mathematical model analysis and numerical simulation for codynamics of meningitis and pneumonia infection with intervention*, Sci. Rep., 12 (2022), 2639.

[20] Martcheva, M. and Crispino-O'Connell, G. *The transmission of meningococcal infection: A mathematical study*, J. Math. Anal. Appl., 283 (2003), 251–275.

[21] Mbaeyi, S.A., Bozio, C.H., Duffy, J., Rubin, L.G., Hariri, S., Stephens, D.S. and MacNeil, J.R. *Meningococcal vaccination: Recommendations of the Advisory Committee on Immunization Practices, United States, 2020*, MMWR Recomm. Rep., 69(9) (2020), 1–41.

[22] Muzumdar, S., Rothe, M.J. and Grant-Kels, J.M. *The rash with maculopapules and fever in adults*, Clin. Dermatol., 37(2) (2019), 109–118.

[23] Ni, H., Knight, A.I., McFadden, J., Cartwright, K. and Palmer, W.H. *Polymerase chain reaction for diagnosis of meningococcal meningitis*, The Lancet, 340(8833) (1992), 1432–1434.

[24] Parikh, S.R., Campbell, H., Bettinger, J.A., Harrison, L.H., Marshall, H.S., Martinon-Torres, F., Safadi, M.A., Shao, Z., Zhu, B., von Gottberg, A., Borrow, R., Ramsay, M.E. and Ladhani, S.N. *The everchanging epidemiology of meningococcal disease worldwide and the potential for prevention through vaccination*, J. Infect., 81(4) (2020), 483–498.

[25] Pizza, M., Bekkat-Berkani, R. and Rappuoli, R. *Vaccines against meningococcal diseases*, Microorganisms, 8(10) (2020), 1521.

[26] Van den, D.P. and Watmough, J. *Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission*, Math. Biosci., 180 (2002), 29–48.

[27] Sorek, N., Ashkenazi, S., Livni, G. and Ben-Zvi, H. *Neisseria meningitidis and cytomegalovirus simultaneous detection in the FilmArray meningitis/encephalitis panel and its clinical relevance*, IDCases, 17 (2019), e00516.

[28] Tilahun, G.T. *Modeling co-dynamics of pneumonia and meningitis diseases*, Adv. Differ. Equ., 2019 (2019), 149.

[29] Trotter, C.L. *Epidemiology of meningococcal disease and vaccine implications*, Vaccine, 29(Suppl. 2) (2010), B18–B25.

[30] United Nations. *World Population Prospects 2024: Demographic Indicators*, Department of Economic and Social Affairs, Population Division, 2024.

[31] Vereen, K. *Transmission dynamics and control strategies of meningococcal infections*, Research Report, 2008.

[32] Workineh, H. and Kassa, S.M. *Optimal control of the spread of meningitis: In the presence of behaviour change of the society and information dependent vaccination*, Commun. Math. Biol. Neurosci. (2021), Article ID 29.

[33] World Health Organization. *Meningococcal Meningitis: Fact Sheet and Surveillance Report*, WHO Press, Geneva, Switzerland, 2010.

[34] Yano, T.K., Bitok, J. and Jerop, R. *Optimal control analysis of meningococcal meningitis disease with varying population size*, Appl. Comput. Math., 11(5) (2022), 140–149.

[35] Zemmal, M.B., Bouaziz, K., Kekana, M.C and Jeeva, N. *Global dynamics at the critical threshold $R_0 = 1$ in a delayed computer virus model with saturated incidence*, J. Math., 2026(1) (2026), Article 5439604.

Iranian Journal of Numerical Analysis and Optimization
Vol. 17, No. 1, 2027, pp 217–246
https://doi.org/10.22067/ijnao.2026.97683.1832
https://ijnao.um.ac.ir/

CC BY NC

Research Article

A predator-prey model incorporating fear effect, prey vigilance costs, and predator hunting cooperation: Analysis and simulations

A. Akinsanya*,

Abstract

In this study, we propose and analyze a predator-prey model that incorporates prey vigilance, fear effects, and predator cooperative hunting to capture key behavioral mechanisms influencing population dynamics. The model explicitly accounts for nonconsumptive effects of predation risk, whereby prey fear and vigilance reduce effective prey growth and modify predator-prey interactions even in the absence of direct predation. We first establish the mathematical structure of the model and identify all biologically meaningful equilibria, including the extinction equilibrium, the predator-free equilibrium, and the coexistence equilibrium. Local stability analysis reveals that fear and vigilance significantly alter classical stability thresholds by reducing prey reproductive potential and limiting predator access to prey. Using Lyapunov-based methods, we derive sufficient conditions for the global asymptotic stability of the extinction and predator-free equilibria, demonstrating that strong fear effects combined with prey vigilance or elevated predator mortality can suppress predator persistence regardless of initial population densities. For the coexistence equilibrium, we show that stable coexistence occurs under balanced levels of fear, vigilance, and predator cooperation, while excessive cooperation destabilizes the equilibrium. This destabilization is characterized analytically by a Hopf bifurcation with respect to the predator cooperation parameter, leading to the emergence of

*Corresponding author

Received 3 February 2026; revised 21 August 2026; accepted 23 August 2026

Ayodeji Akinsanya
Department of Mathematics, University of Tennessee, Knoxville, USA. e-mail: akinsanyaayodeji@gmail.com

How to cite this article
Akinsanya, A., A predator-prey model incorporating fear effect, prey vigilance costs, and predator hunting cooperation: Analysis and simulations. *Iran. J. Numer. Anal. Optim.*, 2027; 17(1): 217-246. https://doi.org/10.22067/ijnao.2026.97683.1832

sustained population oscillations. Numerical simulations, including phase-plane portraits and time-series analyses, are presented to support the theoretical results and confirm the analytical stability and bifurcation findings.

AMS subject classifications (2020): Primary 92D25; Secondary 34D20, 34C23, 37N25.

Keywords: Predator-prey model, Prey vigilance (fear effect), Predator cooperation, Global stability, Hopf bifurcation.

1 Introduction

Classical predator-prey models, such as the Lotka–Volterra equations, assume that predators impact prey solely through direct consumption, with prey growth unaffected by predator presence aside from mortality losses [12, 23]. In reality, however, prey species have evolved a variety of anti-predator behaviors and life-history adjustments in response to predation risk, forcing them to balance the conflicting demands of foraging and avoiding predators [11, 10]. In recent decades, ecologists have recognized that predators can influence prey not only by killing individuals but also by inducing fear and associated nonconsumptive effects on prey physiology and behavior [3, 4, 10]. The mere presence of a predator (even if no kill occurs) can trigger costly defensive responses in prey —such as reduced foraging, heightened stress, or habitat shift— which in turn may lower prey survival or reproduction [3, 4]. These so-called "fear effects" on prey populations can be powerful. For instance, field experiments have demonstrated that chronic predation risk alone (without any actual predation) can significantly depress prey reproductive success [6, 15]. In one landmark study, song sparrows exposed to predator cues (playback of predator sounds) produced about 40% fewer offspring over a breeding season compared to controls, despite suffering no direct predation [6]. Similarly, under high perceived predation risk, large herbivores like elk exhibit reduced fertility and altered physiology [4]. The cascading consequences of fear have been observed at the community level as well —for example, the mere fear of large carnivores can cause trophic cascades by altering the behavior of mesopredators and herbivores [20]. Recent empirical work is now quantifying these demographic impacts of fear in nature, confirming that risk-induced behavioral changes in prey can substantially affect prey population growth [15].

One of the most common prey anti-predator responses is vigilance: Prey invest time and energy in scanning their surroundings for predators, which improves early detection and escape chances [6]. Increased vigilance by an individual can significantly reduce its immediate risk of predation, especially when a predator is nearby [16]. However, time spent being vigilant is time

not spent feeding, courting, or caring for young. Thus, heightened vigilance carries a direct cost in lost foraging opportunities and energy intake [8, 9]. Empirical studies have quantified these costs; for example, foraging efficiency in ungulates declines with increasing vigilance, leading to lower energy gain and potentially reduced fecundity [8, 9]. In the present study, we represent these vigilance costs primarily through reduced prey reproduction rather than through an increase in baseline mortality, reflecting the dominant opportunity-cost mechanism emphasized in the ecological literature. Because vigilance is costly yet beneficial for safety, prey tend to modulate their vigilance level according to the level of predation threat [16]. In periods or areas of higher perceived risk, animals increase their vigilance, whereas they can afford to feed more and scan less under low risk. Social grouping further mediates this trade-off via the "many-eyes" effect [7]: In larger groups, individuals can collectively maintain high overall surveillance while each individual spends less time being vigilant. As a result, per-capita vigilance typically decreases as group size increases, improving the feeding rate of group members [7]. This group-size effect on vigilance has clear fitness benefits —for instance, studies on birds have shown that larger coveys or flocks suffer lower per-capita predation because each individual can devote more time to foraging while still achieving early predator detection in the group [25]. In short, prey vigilance is a dynamic trait that improves survival at the cost of reproduction, and its optimal level reflects a trade-off between safety and feeding [11, 10, 8, 9].

Predators, on the other hand, often benefit from cooperative hunting strategies. Many social predators hunt in groups to subdue prey that would be difficult or dangerous to kill alone [13, 2, 17, 19, 18, 5]. Cooperative hunting has been documented across a wide range of species —from lions and cheetahs to canids like wolves and African wild dogs, and even in some primates [13, 2, 17, 19, 18, 5]. Group hunting frequently leads to higher success rates and allows predators to tackle larger or more formidable prey. For example, the hunting success of African wild dogs increases with pack size [5], and grey wolves in larger packs achieve greater per-capita food intake than lone wolves or small packs [18]. Chimpanzees have also been observed to coordinate hunts of monkeys, increasing their success compared to individual foraging [2]. These observations underscore the evolutionary advantages of group foraging [13]: Larger groups can overwhelm prey defenses and flush out prey more effectively, yielding greater food rewards per predator. An important consequence of cooperative predation is the potential for an Allee effect in predator populations [22]. When predator numbers are very low, hunting success can be so depressed (due to lack of cooperation) that the predators struggle to obtain enough food to sustain themselves or reproduce. But as predator group size grows, hunting efficiency improves nonlinearly, leading to disproportionately higher gains for each additional hunter [5, 18]. Theoretical models of predator-prey dynamics have incorporated such cooperative effects via modified functional responses, which indeed often produce dynamics consistent with an Allee effect [22, 1]. Cooperative hunting

can also introduce destabilizing feedbacks into predator-prey systems, potentially leading to complex dynamics like multiple equilibria or oscillations when predator cooperation is strong [14, 1].

Given the growing evidence that both prey risk-avoidance behaviors and predator cooperative tactics can fundamentally alter ecological outcomes, recent research has focused on extending predator-prey models to include these effects. In particular, much attention has been paid to modeling the fear effect —the impact of predation risk on prey demographics. The incorporation of a fear factor into prey population equations was pioneered by Wang et al. [24], who added a term to represent the reduction in prey birth rate due to anti-predator stress or reduced feeding in the presence of predators. Their mathematical analysis showed that sufficiently high levels of fear can stabilize the predator-prey system by damping or eliminating oscillatory cycles. Subsequent studies have considered fear effects in various contexts and extensions of the basic model [14], consistently finding that antipredator responses tend to have a net stabilizing influence on predator-prey dynamics. On the other hand, predator hunting cooperation has been modeled by modifying the predator's functional response to increase efficiency at higher predator densities [22, 1]. Such models predict that predator cooperation can create bistability or a threshold phenomenon [22]. More recently, Pal et al. [14] analyzed a predator-prey model that included both a fear effect on the prey and cooperative hunting among the predators. Likewise, a very recent study by Takyi et al. [21] examined a model where prey exhibit vigilance behavior and predators engage in cooperative hunting. Their results showed that intense predator cooperation can lead to Hopf bifurcations and bistability, implying oscillatory predator-prey dynamics or sudden regime shifts under high cooperation strength [21]. Overall, the existing modeling literature suggests that prey anti-predator adaptations generally have a stabilizing or buffering effect on predator-prey interactions [24, 14, 21], whereas predator cooperative hunting tends to have a destabilizing effect if cooperation is sufficiently strong [14, 22, 1]. However, to our knowledge, no single theoretical model has yet incorporated all three of these factors —fear, vigilance, and cooperation— together in one framework.

In this paper, we develop and analyze a new predator-prey model that explicitly includes (i) a fear effect in the prey, (ii) prey vigilance behavior, and (iii) predator hunting cooperation. By integrating all three mechanisms, our model provides a more comprehensive representation of predator-prey dynamics under behaviorally realistic assumptions. We derive conditions for the existence of various equilibria and investigate their stability, and we carry out bifurcation analysis to determine how fear, vigilance, and cooperation influence oscillatory dynamics or alternative stable states. Numerical simulations are also performed to illustrate the system's behavior under different parameter regimes.

The remainder of the paper is organized as follows: Section 2 presents the formulation of the mathematical model and defines the key parameters. In Section 3, we explain the boundedness property of the model, while in Section 4 we analyze the equilibrium points of the system and their local and global stability, and we examine possible bifurcations that arise as parameters vary. Section 5 contains numerical simulations that validate and complement the analytical results. Finally, Section 6 concludes the paper with a summary of findings and a discussion of their ecological implications.

2 Model formulation

We consider an unstructured predator-prey system in which $x(t)$ and $y(t)$ denote the prey and predator population densities, respectively, at time t. The prey is assumed to follow logistic growth in the absence of predation, and the predator population increases by consuming prey. In addition to these classical assumptions, we incorporate (i) predator-induced fear, (ii) prey vigilance behaviour, and (iii) cooperative hunting in predators.

We begin with the prey dynamics in the absence of predators, fear, and vigilance. The logistic growth of prey can be separated into three parts: A birth (production) term, a natural death term, and a density-dependent death term due to intra-species competition. Hence,

$$\frac{dx}{dt} = rx - \delta x - \frac{r}{K}x^2,$$

where r denotes the intrinsic birth (production) rate of the prey, δ is the natural death rate, and $\frac{r}{K}x^2$ represents the density-dependent death due to intra-species competition (with K being the carrying capacity).

Next, we incorporate the cost of vigilance. Let $v \in [0, 1]$ denote the level of prey vigilance, with larger values of v corresponding to stronger vigilance. Biologically, vigilance imposes an opportunity cost: Time and energy devoted to scanning for predators are diverted from feeding, mating, and parental care, thereby reducing reproductive output. In this model, we therefore assume that the principal demographic effect of vigilance acts through the prey birth term, while its direct effect on the natural mortality rate δ is negligible or secondary. This assumption is adopted both to isolate the dominant reproductive cost of vigilance and to maintain analytical tractability. It is also consistent with fear-effect models in which predation risk is commonly represented through reduced prey reproduction rather than increased background mortality [24], as well as with empirical studies documenting foraging–vigilance trade-offs and their consequences

for energy intake and fecundity [8, 9]. Accordingly, we modify only the prey production term by multiplying it by $(1 - v)$, while leaving the density-dependent competition term unchanged.

We now incorporate predator-induced fear. Predator presence can suppress prey production through fear-related behavioral responses. To capture this mechanism, we introduce a fear function $f(k, y)$ that depends on predator density y and a fear level parameter $k > 0$, and we again let it act only on the prey production term. In particular, we adopt

$$f(k, y) = \frac{1}{1 + ky}, \qquad k > 0,$$

which is chosen for its simplicity, analytical tractability, and its ability to capture a natural saturation of fear effects as predator density increases. Other functional forms, such as e^{-ky}, are also biologically plausible, but the reciprocal form provides a convenient normalized representation for the present analysis. The sensitivity of the system to alternative fear functions could be explored in future work. This function satisfies the following biologically reasonable properties:

$$\frac{dx}{dt} = r(1 - v)x - \delta x - \frac{r}{K}x^2,$$

$$f(0, y) = 1, \quad f(k, 0) = 1, \quad \lim_{k\to\infty} f(k, y) = 0, \quad \lim_{y\to\infty} f(k, y) = 0,$$

$$\frac{\partial f}{\partial k} < 0, \quad \frac{\partial f}{\partial y} < 0.$$

Accordingly, incorporating both vigilance and fear into the prey production term gives

$$\frac{dx}{dt} = r(1 - v)f(k, y)x - \delta x - \frac{r}{K}x^2.$$

We proceed by introducing predation using a Holling type II functional response. We assume that prey vigilance reduces predation success through a protective factor $(1 + \mu v)$, where $\mu > 0$ measures the effectiveness of vigilance in lowering capture probability. Moreover, we assume that predators cooperate while hunting, so that the effective encounter/attack pressure increases with predator density. Hence, we define the predation (consumption) term as

$$G(x, y, v) = \frac{(q + cy)\, x\, y}{(1 + hx)(1 + \mu v)},$$

where $q > 0$ is the baseline encounter rate, $c \geq 0$ measures the strength of predator hunting cooperation, and $h > 0$ is the Holling type II saturation (handling) parameter. Incorporating predation, the prey equation becomes

$$\frac{dx}{dt} = r(1-v)f(k,y)x - \delta x - \frac{r}{K}x^2 - G(x,y,v).$$

Finally, the predator population increases through consumption of prey with conversion efficiency $\gamma > 0$ and decreases due to natural mortality at rate $\delta_1 > 0$. Thus,

$$\frac{dy}{dt} = \gamma\, G(x,y,v) - \delta_1 y.$$

Combining the above components, we obtain the following nonlinear predator-prey system with fear effect, prey vigilance, and predator hunting cooperation:

$$\begin{aligned} \frac{dx}{dt} &= \frac{r(1-v)\,x}{1+ky} \;-\; \delta x \;-\; \frac{r}{K}x^2 \;-\; \frac{(q+cy)\,x\,y}{(1+hx)(1+\mu v)}, \\ \frac{dy}{dt} &= \gamma\,\frac{(q+cy)\,x\,y}{(1+hx)(1+\mu v)} \;-\; \delta_1 y, \end{aligned} \tag{1}$$

The biological interpretations of all model variables and parameters are summarized in Table 1. We assume positive initial conditions $x(0) = x_0 > 0$ and $y(0) = y_0 > 0$. We assume all parameters are positive unless otherwise stated, and $v \in [0,1]$.

Table 1: Parameters used in the combined predator-prey model.

Parameter	**Description**
x	prey population density
y	predator population density
r	prey birth/production rate
K	prey carrying capacity (controls intra-species competition via $\frac{r}{K}x^2$)
δ	prey natural death rate
δ_1	predator natural death rate
v	level of prey vigilance ($v \in [0,1]$)
k	fear-effect strength (predator-induced suppression of prey production)
μ	effectiveness of vigilance in reducing predation
q	baseline predator encounter/attack rate
c	hunting cooperation strength in predators
h	Holling type II saturation/handling parameter
γ	conversion efficiency (prey biomass to predator biomass)

3 Boundedness of solutions

The boundedness property guarantees that neither population can grow unboundedly due to limited resources and habitat constraints.

Theorem 1. Assume that all parameters of the system are positive and that $v \in [0,1]$. Then every solution $(x(t), y(t))$ of system (1) with initial condition $(x(0), y(0)) \in \mathbb{R}^2_+$ is bounded. In particular, there exists a constant $M > 0$ such that

$$0 \le x(t) \le (1-v)K, \qquad 0 \le x(t) + \frac{1}{\gamma} y(t) \le M \quad \text{for all sufficiently large } t.$$

Proof. First, since

$$\frac{r(1-v)x}{1+ky} \le r(1-v)x \quad \text{and} \quad -\delta x - \frac{(q+cy)xy}{(1+hx)(1+\mu v)} \le 0,$$

we obtain

$$\frac{dx}{dt} \le r(1-v)x - \frac{r}{K}x^2 = rx\left((1-v) - \frac{x}{K}\right).$$

By the comparison principle for scalar differential equations, it follows that

$$\limsup_{t\to\infty} x(t) \le (1-v)K.$$

Next, define the auxiliary function

$$W(t) = x(t) + \frac{1}{\gamma} y(t).$$

Differentiating and using the system equations yields

$$\frac{dW}{dt} = \frac{dx}{dt} + \frac{1}{\gamma}\frac{dy}{dt} = \frac{r(1-v)x}{1+ky} - \delta x - \frac{r}{K}x^2 - \frac{\delta_1}{\gamma} y.$$

Dropping the nonpositive term $-\frac{r}{K}x^2$ and using $\frac{r(1-v)x}{1+ky} \le r(1-v)x$, we obtain

$$\frac{dW}{dt} \le r(1-v)x - \delta x - \frac{\delta_1}{\gamma} y \le r(1-v)x - \delta x - \frac{\delta_1}{\gamma} y.$$

Using the bound $\limsup_{t\to\infty} x(t) \le (1-v)K$, we have for sufficiently large t,

$$r(1-v)x(t) \le r(1-v)^2 K.$$

Hence, for large t,

$$\frac{dW}{dt} \leq r(1-v)^2K - \delta x - \frac{\delta_1}{\gamma}y.$$

Let $\varepsilon = \min\{\delta, \delta_1\} > 0$. Then

$$-\delta x - \frac{\delta_1}{\gamma}y \leq -\varepsilon\left(x + \frac{1}{\gamma}y\right) = -\varepsilon W(t),$$

so that for large t,

$$\frac{dW}{dt} + \varepsilon W(t) \leq r(1-v)^2K.$$

Solving this differential inequality gives

$$\limsup_{t\to\infty} W(t) \leq \frac{r(1-v)^2K}{\varepsilon}.$$

Therefore, both $x(t)$ and $y(t)$ are bounded for $t \geq 0$, and all solutions initiating from $\mathbb{R}^2_+$ are bounded. □

4 Equilibria

To obtain the equilibria of the system, we solve $\dot{x} = 0$ and $\dot{y} = 0$ simultaneously for system (1). The system possesses the following nonnegative equilibria:

1. Extinction Equilibrium: $E_0 = (0, 0)$.

2. Predator-free equilibrium: $E_1 = (x_1, 0)$, where

$$x_1 = K\left(1 - v - \frac{\delta}{r}\right).$$

   We note that E_1 is feasible whenever $x_1 > 0$, equivalently, if $r(1-v) > \delta$.

3. Coexistence equilibrium: $E_2 = (x^*, y^*)$, where (x^*, y^*) is a coexistence equilibrium with $x^* > 0$ and $y^* > 0$.

To derive the coexistence equilibrium, set $\dot{y} = 0$. If $y^* > 0$, then

$$0 = \gamma\frac{(q + cy^*)x^*y^*}{(1 + hx^*)(1 + \mu v)} - \delta_1 y^* \quad \Longrightarrow \quad \gamma\frac{(q + cy^*)x^*}{(1 + hx^*)(1 + \mu v)} = \delta_1.$$

Define

$$A = \frac{\gamma}{1+\mu v} > 0.$$

Then the above relation becomes

$$q + cy^* = \frac{\delta_1(1+hx^*)}{Ax^*},$$

and hence

$$y^* = \frac{1}{c}\left(\frac{\delta_1(1+hx^*)}{Ax^*} - q\right).$$

Therefore, E_2 is feasible only if $y^* > 0$, that is,

$$\frac{\delta_1(1+hx^*)}{Ax^*} > q \quad \Longleftrightarrow \quad 0 < x^* < \frac{\delta_1}{Aq - h\delta_1},$$

which requires $Aq - h\delta_1 > 0$.

Next, set $\dot{x} = 0$ and assume $x^* > 0$. Dividing the prey equation by x^* yields

$$0 = \frac{r(1-v)}{1+ky^*} - \delta - \frac{r}{K}x^* - \frac{(q+cy^*)y^*}{(1+hx^*)(1+\mu v)}.$$

Using $q + cy^* = \dfrac{\delta_1(1+hx^*)}{Ax^*}$ and $A = \dfrac{\gamma}{1+\mu v}$, we obtain

$$\frac{(q+cy^*)y^*}{(1+hx^*)(1+\mu v)} = \frac{\delta_1 y^*}{\gamma x^*}.$$

Hence, the coexistence equilibrium satisfies

$$\frac{r(1-v)}{1+ky^*} = \delta + \frac{r}{K}x^* + \frac{\delta_1}{\gamma}\frac{y^*}{x^*},$$

together with

$$y^* = \frac{1}{c}\left(\frac{\delta_1(1+hx^*)}{Ax^*} - q\right), \qquad A = \frac{\gamma}{1+\mu v}.$$

Once a positive real root x^* satisfying the feasibility conditions is obtained, y^* follows uniquely from the above expression, and thus $E_2 = (x^*, y^*)$.

The stability of the equilibrium points is determined by the nature of the eigenvalues of the Jacobian matrix $J(x, y)$ of system (1).

Let

$$F(x,y) = \frac{r(1-v)x}{1+ky} - \delta x - \frac{r}{K}x^2 - \frac{(q+cy)xy}{(1+hx)(1+\mu v)},$$

$$H(x,y) = \gamma \frac{(q+cy)xy}{(1+hx)(1+\mu v)} - \delta_1 y.$$

Then

$$J(x,y) = \begin{pmatrix} \frac{\partial F}{\partial x} & \frac{\partial F}{\partial y} \\ \frac{\partial H}{\partial x} & \frac{\partial H}{\partial y} \end{pmatrix}.$$

A direct computation gives

$$J(x,y) = \begin{pmatrix} \frac{r(1-v)}{1+ky} - \delta - \frac{2r}{K}x - \frac{y(q+cy)}{(1+\mu v)(1+hx)^2} & -\frac{rk(1-v)x}{(1+ky)^2} - \frac{x(q+2cy)}{(1+\mu v)(1+hx)} \\ \gamma \frac{y(q+cy)}{(1+\mu v)(1+hx)^2} & \gamma \frac{x(q+2cy)}{(1+\mu v)(1+hx)} - \delta_1 \end{pmatrix}.$$

Theorem 2. The extinction equilibrium point $E_0 = (0,0)$ of system (1) is locally asymptotically stable if $r(1-v) < \delta$, and it is unstable (a saddle point) if $r(1-v) > \delta$.

Proof. Evaluating the Jacobian $J(x,y)$ at $E_0 = (0,0)$ yields

$$J_{E_0} = J(0,0) = \begin{pmatrix} r(1-v) - \delta & 0 \\ 0 & -\delta_1 \end{pmatrix}.$$

Therefore, the eigenvalues of $J(E_0)$ are

$$\lambda_1 = r(1-v) - \delta, \qquad \lambda_2 = -\delta_1.$$

Since $\delta_1 > 0$, we have $\lambda_2 < 0$. Therefore, both eigenvalues have negative real parts if and only if $r(1-v) - \delta < 0$, that is, $r(1-v) < \delta$, and in this case E_0 is locally asymptotically stable. If $r(1-v) > \delta$, then $\lambda_1 > 0$ and $\lambda_2 < 0$, so E_0 is a saddle point and hence unstable. □

Theorem 3. Assume that $r(1-v) > \delta$ so that the predator-free equilibrium point

$$E_1 = (x_1, 0), \qquad x_1 = K\left(1 - v - \frac{\delta}{r}\right),$$

is feasible. Then E_1 is locally asymptotically stable if

$$\gamma \frac{qx_1}{(1+\mu v)(1+hx_1)} < \delta_1,$$

and it is unstable (a saddle point) if

$$\gamma \frac{qx_1}{(1+\mu v)(1+hx_1)} > \delta_1.$$

Proof. Evaluating the Jacobian $J(x,y)$ at $E_1 = (x_1, 0)$ yields

$$J_{E_1} = J(x_1, 0) = \begin{pmatrix} r(1-v) - \delta - \frac{2r}{K}x_1 & -rk(1-v)x_1 - \frac{qx_1}{(1+\mu v)(1+hx_1)} \\ 0 & \gamma \frac{qx_1}{(1+\mu v)(1+hx_1)} - \delta_1 \end{pmatrix}.$$

Since $x_1 = K\left(1 - v - \frac{\delta}{r}\right)$, we obtain

$$r(1-v) - \delta - \frac{2r}{K}x_1 = -\big(r(1-v) - \delta\big).$$

Hence,

$$J_{E_1} = \begin{pmatrix} -\big(r(1-v) - \delta\big) & * \\ 0 & \gamma \frac{qx_1}{(1+\mu v)(1+hx_1)} - \delta_1 \end{pmatrix},$$

which is an upper triangular matrix. Therefore, the eigenvalues of $J(E_1)$ are

$$\lambda_1 = -\big(r(1-v) - \delta\big), \qquad \lambda_2 = \gamma \frac{qx_1}{(1+\mu v)(1+hx_1)} - \delta_1.$$

Since $r(1-v) > \delta$, we have $\lambda_1 < 0$. Thus, E_1 is locally asymptotically stable if and only if $\lambda_2 < 0$, that is,

$$\gamma \frac{qx_1}{(1+\mu v)(1+hx_1)} < \delta_1.$$

If $\gamma \frac{qx_1}{(1+\mu v)(1+hx_1)} > \delta_1$, then $\lambda_2 > 0$ while $\lambda_1 < 0$, so E_1 is a saddle point and hence unstable. □

Theorem 4. The coexistence equilibrium $E_2 = (x^*, y^*)$ with $x^* > 0$ and $y^* > 0$ is locally asymptotically stable if

$$\operatorname{tr}\big(J(E_2)\big) < 0 \qquad \text{and} \qquad \det\big(J(E_2)\big) > 0,$$

and it is unstable if either $\operatorname{tr}\big(J(E_2)\big) > 0$ or $\det\big(J(E_2)\big) < 0$.

Proof. Evaluating the Jacobian matrix $J(x,y)$ at $E_2 = (x^*, y^*)$ yields

$$J(E_2) = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},$$

where

$$a_{11} = \frac{r(1-v)}{1+ky^*} - \delta - \frac{2r}{K}x^* - \frac{y^*(q+cy^*)}{(1+\mu v)(1+hx^*)^2},$$

$$a_{12} = -\frac{rk(1-v)x^*}{(1+ky^*)^2} - \frac{x^*(q+2cy^*)}{(1+\mu v)(1+hx^*)},$$

$$a_{21} = \gamma\,\frac{y^*(q+cy^*)}{(1+\mu v)(1+hx^*)^2},$$

$$a_{22} = \gamma\,\frac{x^*(q+2cy^*)}{(1+\mu v)(1+hx^*)} - \delta_1.$$

The characteristic equation of $J(E_2)$ is

$$\det(\lambda I - J(E_2)) = 0 \quad \Longleftrightarrow \quad \lambda^2 - \operatorname{tr}\big(J(E_2)\big)\lambda + \det\big(J(E_2)\big) = 0,$$

where

$$\operatorname{tr}\big(J(E_2)\big) = a_{11} + a_{22}, \qquad \det\big(J(E_2)\big) = a_{11}a_{22} - a_{12}a_{21}.$$

By the Routh-Hurwitz criterion for a second-order polynomial, both eigenvalues of $J(E_2)$ have negative real parts if and only if

$$\operatorname{tr}\big(J(E_2)\big) < 0 \quad \text{and} \quad \det\big(J(E_2)\big) > 0.$$

Therefore, E_2 is locally asymptotically stable when the above inequalities hold. If instead $\operatorname{tr}\big(J(E_2)\big) > 0$, then at least one eigenvalue has positive real part and hence E_2 is unstable. If $\det\big(J(E_2)\big) < 0$, then the eigenvalues have opposite signs, implying that E_2 is a saddle point and thus unstable. □

Theorem 5. The extinction equilibrium point $E_0 = (0,0)$ of system (1) is globally asymptotically stable in $\mathbb{R}^2_+$ if

$$r(1-v) < \delta.$$

Proof. Consider the Lyapunov function

$$V(x,y) = x + \frac{1}{\gamma}y.$$

Clearly, $V(0,0) = 0$ and $V(x,y) > 0$ for $(x,y) \neq (0,0)$ in $\mathbb{R}^2_+$. Differentiating V along solutions gives

$$\dot{V} = \dot{x} + \frac{1}{\gamma}\dot{y} = \frac{r(1-v)x}{1+ky} - \delta x - \frac{r}{K}x^2 - \frac{\delta_1}{\gamma}y \leq (r(1-v)-\delta)x - \frac{r}{K}x^2 - \frac{\delta_1}{\gamma}y.$$

If $r(1-v) < \delta$, then $(r(1-v)-\delta)x \leq 0$ for $x \geq 0$, hence

$$\dot{V} \leq -\frac{r}{K}x^2 - \frac{\delta_1}{\gamma}y < 0 \qquad \text{for all } (x,y) \neq (0,0).$$

Therefore, V is strictly decreasing along all nontrivial solutions, and by Lyapunov's direct method, $E_0 = (0,0)$ is globally asymptotically stable in $\mathbb{R}^2_+$. □

Theorem 6. Assume that $r(1-v) > \delta$ so that the predator-free equilibrium

$$E_1 = (x_1, 0), \qquad x_1 = K\left(1 - v - \frac{\delta}{r}\right),$$

is feasible. Let

$$\varepsilon = \min\{\delta, \delta_1\}.$$

From the boundedness result (Theorem 1), we have

$$\limsup_{t\to\infty} W(t) \le \frac{r(1-v)^2 K}{\varepsilon},$$

where $W(t) = x(t) + \frac{1}{\gamma}y(t)$. Since $y(t) \le \gamma W(t)$ for all $t \ge 0$, it follows that

$$\limsup_{t\to\infty} y(t) \le \chi, \qquad \text{where} \qquad \chi = \frac{\gamma r(1-v)^2 K}{\varepsilon}.$$

If

$$1 < \frac{(1+\mu v)\delta_1}{\gamma x_1 (q + c\chi)},$$

then the predator-free equilibrium E_1 is globally asymptotically stable in the region

$$\Omega = \{(x,y) \in \mathbb{R}^2_+ : \ x > 0, \ y \ge 0\}.$$

Proof. Consider the Lyapunov function

$$V(x,y) = A_1\left[x - x_1 - x_1 \ln\left(\frac{x}{x_1}\right) + \frac{y}{\gamma}\right],$$

where $A_1 > 0$ is a constant to be chosen. Clearly, $V(x_1, 0) = 0$ and $V(x,y) > 0$ for all $(x,y) \ne (x_1, 0)$ with $x > 0$, $y \ge 0$. Differentiating V along solutions yields

$$\dot{V} = A_1\left[\left(1 - \frac{x_1}{x}\right)\dot{x} + \frac{1}{\gamma}\dot{y}\right].$$

Substituting $\dot{x}$ and $\dot{y}$ and simplifying the predation terms gives

$$\dot{V} = A_1\left(\left(1 - \frac{x_1}{x}\right)\left[\frac{r(1-v)x}{1+ky} - \delta x - \frac{r}{K}x^2\right] + \frac{(q+cy)x_1 y}{(1+hx)(1+\mu v)} - \frac{\delta_1}{\gamma}y\right).$$

Since $\frac{1}{1+ky} \leq 1$, we have

$$\left(1-\frac{x_1}{x}\right)\left[\frac{r(1-v)x}{1+ky}-\delta x-\frac{r}{K}x^2\right] \leq \left(1-\frac{x_1}{x}\right)\left[r(1-v)x-\delta x-\frac{r}{K}x^2\right] = -\frac{r}{K}(x-x_1)^2 \leq 0,$$

where we used $r(1-v)-\delta = \frac{r}{K}x_1$. Therefore,

$$\dot{V} \leq A_1\left[-\frac{r}{K}(x-x_1)^2 + y\left(\frac{(q+cy)x_1}{(1+hx)(1+\mu v)}-\frac{\delta_1}{\gamma}\right)\right].$$

Using $(1+hx) \geq 1$ and the ultimate bound $y(t) \leq \chi$ (for large t), we obtain for large t,

$$\frac{(q+cy)x_1}{(1+hx)(1+\mu v)} \leq \frac{(q+c\chi)x_1}{1+\mu v}.$$

Hence, for large t,

$$\dot{V} \leq A_1\left[-\frac{r}{K}(x-x_1)^2 + y\left(\frac{(q+c\chi)x_1}{1+\mu v}-\frac{\delta_1}{\gamma}\right)\right].$$

If

$$1 < \frac{(1+\mu v)\delta_1}{\gamma x_1(q+c\chi)} \quad \Longleftrightarrow \quad \frac{(q+c\chi)x_1}{1+\mu v}-\frac{\delta_1}{\gamma} < 0,$$

then $\dot{V} < 0$ for all $(x,y) \neq (x_1,0)$ (for sufficiently large t). Thus, V is a Lyapunov function ensuring global asymptotic stability of E_1 in Ω by Lyapunov's asymptotic stability theorem. □

Remark 1. The sufficient condition in Theorem 6 may still hold even when $\delta_1 < \delta$, although it becomes more restrictive as δ_1 decreases. Indeed, when $\delta_1 < \delta$, we have

$$\varepsilon = \min\{\delta, \delta_1\} = \delta_1, \qquad \chi = \frac{\gamma r(1-v)^2 K}{\delta_1}.$$

For example, let

$$r = 0.8, \quad K = 5, \quad v = 0.1, \quad \delta = 0.20, \quad \delta_1 = 0.15, \quad \mu = 0.5,$$

$$\gamma = 0.2, \quad q = 0.1, \quad c = 0.02.$$

Then

$$x_1 = K\left(1-v-\frac{\delta}{r}\right) = 5\,(0.9-0.25) = 3.25,$$

and

$$\chi = \frac{\gamma r(1-v)^2 K}{\delta_1} = \frac{0.2\cdot 0.8\cdot (0.9)^2\cdot 5}{0.15} = 4.32.$$

Hence,

$$\frac{(1+\mu v)\delta_1}{\gamma x_1(q+c\chi)} = \frac{(1+0.5\cdot 0.1)(0.15)}{0.2\cdot 3.25\cdot(0.1+0.02\cdot 4.32)} \approx 1.30 > 1.$$

Therefore, the sufficient condition in Theorem 6 is satisfied in this case even though $\delta_1 < \delta$.

From an ecological perspective, the quantity χ represents an ultimate upper bound on predator density obtained from the boundedness estimate, rather than the actual long-term predator population level. Thus, the condition in Theorem 6 is sufficient and may be conservative, but it still provides a useful criterion ensuring that predator pressure remains weak enough for the predator-free equilibrium to be globally asymptotically stable.

Theorem 7. Assume that system (1) admits a unique coexistence equilibrium $E_2 = (x^*, y^*)$ with $x^* > 0$ and $y^* > 0$. Further assume that all solutions initiating in $\mathbb{R}^2_+$ are bounded and that the positive quadrant contains a positively invariant compact set

$$\Omega^* = \{(x,y) \in \mathbb{R}^2_+ : \ 0 < x \le X, \ 0 < y \le \chi\},$$

for some constants $X > 0$ and $\chi > 0$.

Suppose additionally that there exists a continuously differentiable Dulac function $B(x,y) > 0$ on Ω^* such that

$$\frac{\partial}{\partial x}\big(B(x,y)\dot{x}\big) + \frac{\partial}{\partial y}\big(B(x,y)\dot{y}\big) < 0 \qquad \text{for all } (x,y) \in \Omega^*.$$

If, moreover, E_2 is locally asymptotically stable (i.e., $\mathrm{tr}(J(E_2)) < 0$ and $\det(J(E_2)) > 0$), then E_2 is globally asymptotically stable in Ω^*.

Proof. By the boundedness result, every solution $(x(t), y(t))$ with initial condition in $\mathbb{R}^2_+$ eventually enters Ω^* and remains there for all sufficiently large t. Hence, the ω-limit set of any solution is nonempty, compact, invariant, and contained in Ω^*.

Since $B(x,y) > 0$ on Ω^* and

$$\frac{\partial}{\partial x}\big(B(x,y)\dot{x}\big) + \frac{\partial}{\partial y}\big(B(x,y)\dot{y}\big) < 0 \quad \text{in } \Omega^*,$$

the Bendixson–Dulac criterion implies that system (1) admits no nontrivial periodic orbits lying in Ω^*.

Because E_2 is the unique equilibrium in the interior of Ω^* and periodic orbits are excluded, the Poincaré-Bendixson theorem implies that the ω-limit set of any interior trajectory must be an equilibrium point. Therefore, every solution in Ω^* converges to E_2.

Finally, since E_2 is locally asymptotically stable, the convergence is to E_2 and not to any boundary equilibrium. Hence E_2 is globally asymptotically stable in Ω^*. □

Theorem 8. Assume that the coexistence equilibrium $E_2 = (x^*, y^*)$ of system (1) exists with $x^* > 0$ and $y^* > 0$. Let J_{E_2} denote the Jacobian matrix of system (1) evaluated at E_2. Then system (1) undergoes a Hopf bifurcation at E_2 with respect to the bifurcation parameter c provided that the following properties hold:

(i) $\operatorname{tr}(J^*_{E_2}) = 0$,

(ii) $\det(J^*_{E_2}) > 0$,

(iii) $\dfrac{d}{dc}\big(\operatorname{tr}(J_{E_2})\big)\Big|_{c=c^*} \neq 0$,

where $J^*_{E_2} = J_{E_2}\big|_{c=c^*}$ and c^* denote the critical value of the predator cooperation parameter c for which the coexistence equilibrium E_2 satisfies $\operatorname{tr}(J_{E_2}) = 0$ and $\det(J_{E_2}) > 0$.

Proof. At $E_2 = (x^*, y^*)$, the equilibrium condition from $\dot{y} = 0$ gives

$$0 = \gamma \frac{(q + cy^*)x^*y^*}{(1 + hx^*)(1 + \mu v)} - \delta_1 y^* \quad \Longrightarrow \quad \gamma \frac{(q + cy^*)x^*}{(1 + hx^*)(1 + \mu v)} = \delta_1.$$

Define

$$A = \frac{\delta_1(1 + hx^*)(1 + \mu v)}{\gamma x^*} > 0.$$

Then the above relation becomes

$$q + cy^* = A \qquad \Longrightarrow \qquad y^* = \frac{A - q}{c},$$

so E_2 is feasible only if $A > q$.

Next, evaluating the Jacobian at E_2 and using $q + cy^* = A$ and $q + 2cy^* = 2A - q$ yields

$$J_{E_2} = \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix},$$

where

$$J_{11} = \frac{r(1 - v)}{1 + ky^*} - \delta - \frac{2r}{K}x^* - \frac{y^*A}{(1 + \mu v)(1 + hx^*)^2},$$

$$J_{12} = -\frac{rk(1 - v)x^*}{(1 + ky^*)^2} - \frac{x^*(2A - q)}{(1 + \mu v)(1 + hx^*)},$$

$$J_{21} = \gamma \frac{y^* A}{(1+\mu v)(1+hx^*)^2},$$

$$J_{22} = \gamma \frac{x^*(2A-q)}{(1+\mu v)(1+hx^*)} - \delta_1 = \delta_1 - \frac{\gamma q x^*}{(1+\mu v)(1+hx^*)}.$$

To obtain a tractable expression for the trace, we also use the prey equilibrium condition $\dot{x} = 0$, which after dividing by $x^* > 0$ and using $q + cy^* = A$ becomes

$$\frac{r(1-v)}{1+ky^*} = \delta + \frac{r}{K}x^* + \frac{Ay^*}{(1+\mu v)(1+hx^*)} = \delta + \frac{r}{K}x^* + \frac{\delta_1 y^*}{\gamma x^*}.$$

Substituting this identity into $\mathrm{tr}(J_{E_2}) = J_{11} + J_{22}$ simplifies the trace to

$$\mathrm{tr}(J_{E_2}) = -\frac{r}{K}x^* + \frac{\delta_1 h\, y^*}{\gamma(1+hx^*)} + \delta_1 - \frac{\gamma q x^*}{(1+\mu v)(1+hx^*)}.$$

Hence, $\mathrm{tr}(J_{E_2}) = 0$ if and only if

$$\frac{\delta_1 h\, y^*}{\gamma(1+hx^*)} = \frac{r}{K}x^* - \delta_1 + \frac{\gamma q x^*}{(1+\mu v)(1+hx^*)}.$$

Using $y^* = \dfrac{A-q}{c}$, we obtain the critical value

$$c = c^* = \frac{A-q}{\dfrac{\gamma(1+hx^*)}{\delta_1 h}\left(\dfrac{r}{K}x^* - \delta_1 + \dfrac{\gamma q x^*}{(1+\mu v)(1+hx^*)}\right)},$$

$$A = \frac{\delta_1(1+hx^*)(1+\mu v)}{\gamma x^*}.$$

Let $J^*_{E_2} = J_{E_2}\big|_{c=c^*}$. Condition (ii), namely $\det(J^*_{E_2}) > 0$, guarantees that the eigenvalues of $J^*_{E_2}$ are purely imaginary, that is,

$$\lambda_{1,2} = \pm i\sqrt{\det(J^*_{E_2})}.$$

It remains to verify the transversality condition (iii). Since $x^* = x^*(c)$ depends on c through the equilibrium equations, we have

$$\frac{d}{dc}\big(\mathrm{tr}(J_{E_2})\big)\Big|_{c=c^*} = \omega_1 \left.\frac{dx^*}{dc}\right|_{c=c^*} + \omega_2,$$

where

$$\omega_2 = \frac{\partial}{\partial c}\big(\mathrm{tr}(J_{E_2})\big)\Big|_{c=c^*} = -\frac{\delta_1 h}{\gamma(1+hx^*)}\,\frac{A-q}{(c^*)^2},$$

and

$$\omega_1 = \frac{\partial}{\partial x^*}\big(\mathrm{tr}(J_{E_2})\big)\Big|_{c=c^*} = -\frac{r}{K} - \frac{\delta_1^2 h(1+\mu v)}{\gamma^2 c^*(x^*)^2(1+hx^*)} - \frac{\delta_1 h^2 y^*}{\gamma(1+hx^*)^2} - \frac{\gamma q}{(1+\mu v)(1+hx^*)^2},$$

using the identity

$$\frac{dA}{dx^*} = -\frac{\delta_1(1+\mu v)}{\gamma(x^*)^2}, \qquad y^* = \frac{A-q}{c^*}.$$

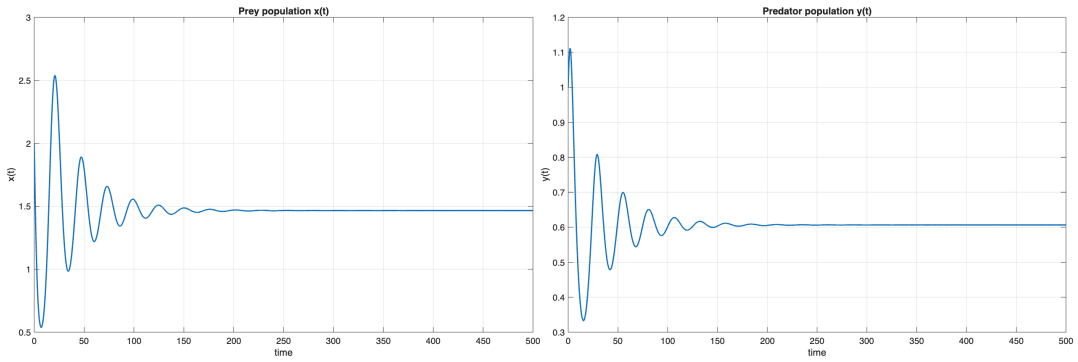
Therefore, if

$$\omega_1 \left.\frac{dx^*}{dc}\right|_{c=c^*} + \omega_2 \neq 0,$$

then the eigenvalues cross the imaginary axis with nonzero speed as c passes through c^*, and by the Hopf bifurcation theorem the system undergoes a Hopf bifurcation at E_2 with respect to c. □

5 Numerical simulations

To gain a deeper understanding of the proposed predator-prey model, we numerically solved the system using the MATLAB `ode45` solver. In the numerical simulations, we selected positive nondimensional parameter values that are illustrative rather than calibrated to any specific empirical predator–prey system. Nevertheless, the chosen values are biologically plausible and lie within ranges commonly adopted in theoretical studies of predator–prey dynamics, while also satisfying the feasibility and stability conditions derived in the analytical sections. The parameter values used for the simulations of Figure 1 are $r = 1$, $K = 10$, $\delta = 0.1766$, $\delta_1 = 0.2429$, $v = 0.2544$, $\mu = 1.3809$, $k = 0.9421$, $q = 0.3411$, $c = 0.1636$, $h = 0.2089$, $\gamma = 0.6643$. The time evolution of the prey population $x(t)$ and the predator population $y(t)$ are depicted in Figures 1a and 1b, respectively. The numerical results show that both populations converge to a positive interior equilibrium point $E_2 = (x^*, y^*) \approx (1.4660, 0.6065)$, indicating the coexistence of the prey and predator species. Moreover, the solution trajectories approach E_2 as time increases, confirming that the coexistence equilibrium is locally asymptotically stable, in agreement with the theoretical stability analysis. Figure 2 illustrates the phase-plane dynamics of the predator-prey system for different combinations of the prey vigilance level v and the predator cooperation parameter c, while all other parameters are held fixed. The parameter values used in all panels are $r = 1$, $K = 10$, $\delta = 0.1766$, $\delta_1 = 0.2429$, $\mu = 1.3809$, $k = 0.9421$, $q = 0.3411$, $h = 0.2089$, $\gamma = 0.6643$, with the values of v and c varying across panels as indicated. The black arrows represent the vector field, the red and green curves denote the prey nullcline $\dot{x} = 0$ and



(a) (b)

Figure 1: Time-series simulation of system (1). Panel (a) shows the prey population $x(t)$ versus time, while panel (b) shows the predator population $y(t)$ versus time.

predator nullcline $\dot{y} = 0$, respectively, the blue curves show solution trajectories initiated from positive initial conditions, and the equilibrium points are marked in yellow.

In panel (a), where the vigilance level is relatively high and predator cooperation is weak, the predator nullcline does not intersect the prey nullcline in the interior of the positive quadrant. As a result, no coexistence equilibrium exists, and all trajectories are directed toward the predator-free equilibrium E_1. This indicates that strong prey vigilance combined with weak cooperative hunting suppresses predator persistence, ultimately leading to predator extinction.

Panels (b) and (c) show that reducing the vigilance level or increasing the cooperation parameter shifts the predator nullcline upward, leading to an intersection with the prey nullcline in the interior of the phase plane. This intersection corresponds to the coexistence equilibrium E_2. The direction of the vector field around E_2, together with the convergence of nearby trajectories, indicates that E_2 is locally asymptotically stable. Biologically, this reflects a balance between prey defense mechanisms and predator cooperative hunting that allows long-term coexistence of both populations.

In panel (d), which corresponds to the baseline parameter set $(v, c) = (0.2544, 0.1636)$, the coexistence equilibrium E_2 is clearly visible and acts as a stable spiral sink. All trajectories starting in the positive quadrant converge to this equilibrium, demonstrating that the system returns to steady coexistence even after moderate perturbations in population densities.

Panels (e) and (f) further emphasize the role of predator cooperation. As the cooperation parameter c increases, the predator nullcline shifts further upward, enhancing predator growth at higher population densities. Although coexistence persists, the orientation of the vector field and the shape of the trajectories indicate increasingly pronounced oscillatory transients before

convergence. This behavior suggests that the system approaches the threshold of more complex dynamics, such as sustained oscillations, as cooperative hunting becomes stronger.

Overall, the phase-plane analysis in Figure 2 demonstrates that prey vigilance and predator cooperation jointly regulate the qualitative dynamics of the system. High vigilance tends to destabilize predator persistence, whereas sufficient cooperative hunting enables predators to overcome prey defenses and coexist stably with the prey population. These numerical observations are fully consistent with the analytical equilibrium and stability results established earlier.

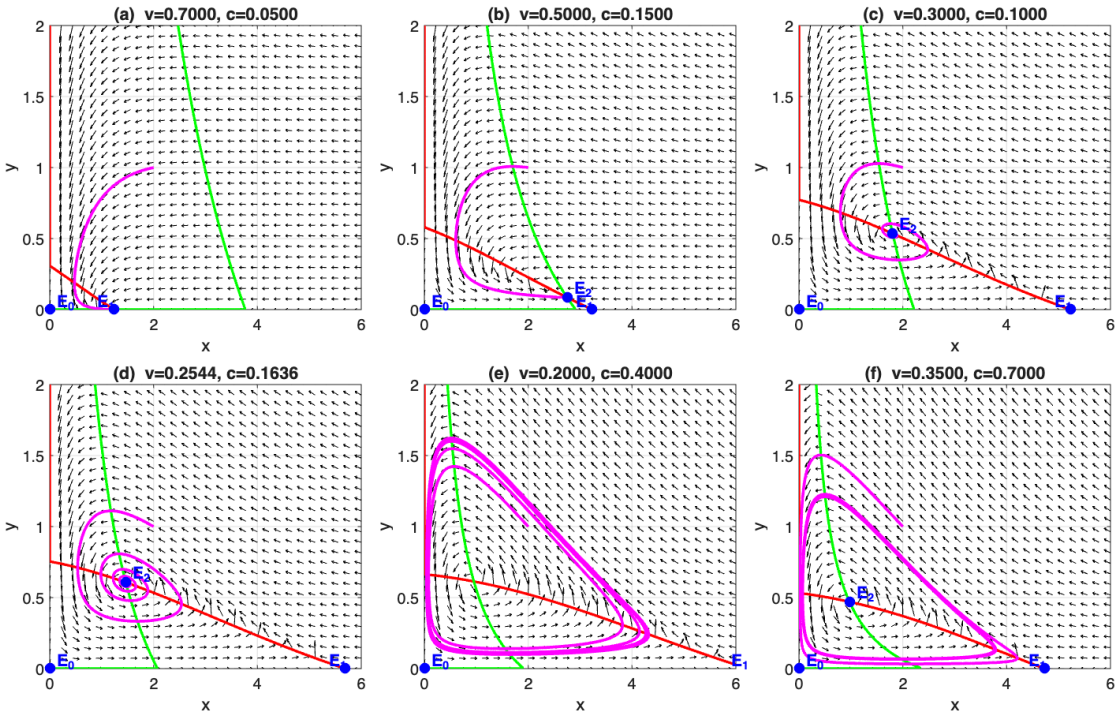


Figure 2: Phase-plane portraits of system (1) for different combinations of prey vigilance v and predator cooperation c.

Figure 3 presents the numerical phase-portrait simulations of the system using the same parameter set, $r = 1.00$, $v = 0.20$, $K = 10$, $\delta = 0.10$, $\delta_1 = 0.10$, $k = 0.50$, $\mu = 0.50$, $q = 0.50$, $c = 0.20$, $h = 0.20$, $\gamma = 0.70$. In Figure 3A, the trajectories illustrate the global behavior of the system under this parameter regime, showing convergence toward the extinction equilibrium $E_0 = (0, 0)$ when the corresponding stability condition is satisfied. Figure 3B depicts the predator-free regime, where the predator-free equilibrium $E_1 = (x_1, 0)$ exists with $x_1 > 0$ and attracts all solution trajectories initiated from positive initial conditions, indicating prey persistence in the absence of predators. Finally, Figure 3C demonstrates the coexistence

regime for the same parameter values, where an interior equilibrium $E_2 = (x^*, y^*)$ exists and all trajectories starting in the positive quadrant converge to E_2, confirming long-term coexistence of prey and predator populations. Overall, these panels numerically corroborate the analytical global stability results derived earlier.

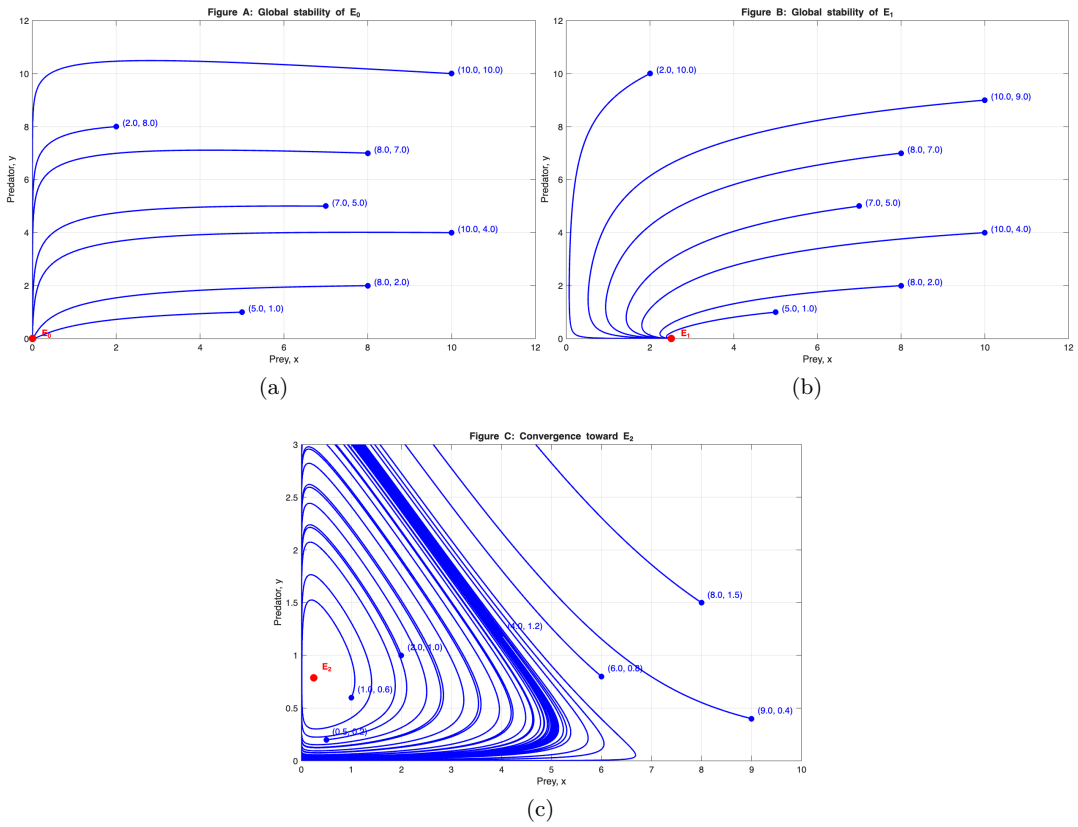


Figure 3: Numerical phase portraits illustrating three qualitative dynamical regimes of system (1). This figure is intended to illustrate the global behaviors predicted by Theorems 5–7.

Figures 4–6 provide numerical evidence of a Hopf bifurcation at the coexistence equilibrium $E_2 = (x^*, y^*)$ of system (1) with respect to the predator cooperation parameter c, in agreement with the analytical conditions stated in Theorem 8. Unless otherwise stated, the parameter values used in all simulations are $r = 1.00$, $v = 0.20$, $K = 10$, $\delta = 0.10$, $\delta_1 = 0.10$, $k = 0.50$, $\mu = 0.50$, $q = 0.50$, $h = 0.20$, $\gamma = 0.70$, while c is treated as the bifurcation parameter.

Figure 4 illustrates the numerical continuation of the predator population extrema as the cooperation parameter c varies. The maximum and minimum values of $y(t)$, computed after

discarding transient dynamics, are plotted against c. For $c < c^*$, both extrema coincide, indicating convergence to the coexistence equilibrium E_2. As c increases and crosses the critical value $c = c^*$, the extrema separate smoothly, signaling the emergence of a stable periodic orbit. This behavior confirms the birth of a limit cycle through a Hopf bifurcation at $c = c^*$.

Figure 5 presents time-series simulations for the prey and predator populations when $c < c^*$. Both $x(t)$ and $y(t)$ exhibit damped oscillations that decay toward constant values, corresponding to the locally asymptotically stable coexistence equilibrium E_2. The associated phase portrait shows inward spiraling trajectories converging to E_2, further confirming its stability below the Hopf threshold.

Figure 6 displays the system dynamics for $c > c^*$. In this regime, the coexistence equilibrium E_2 becomes unstable, and the trajectories converge to a closed orbit in the phase plane. The corresponding time-series exhibit sustained oscillations with constant amplitude, which is characteristic of a stable limit cycle generated by a supercritical Hopf bifurcation. Additionally, the Jacobian trace evaluated at E_2 crosses zero at $c = c^*$ while the determinant remains positive, numerically validating conditions (i) and (ii) of Theorem 8. The smooth and nonzero slope of the trace at c^* confirms the transversality condition $\frac{d}{dc}\mathrm{tr}(J_{E_2})\big|_{c=c^*} \neq 0$.

Overall, Figures 4–6 collectively demonstrate that increasing predator cooperation destabilizes the coexistence equilibrium through a Hopf bifurcation, leading to persistent oscillations in both prey and predator populations.

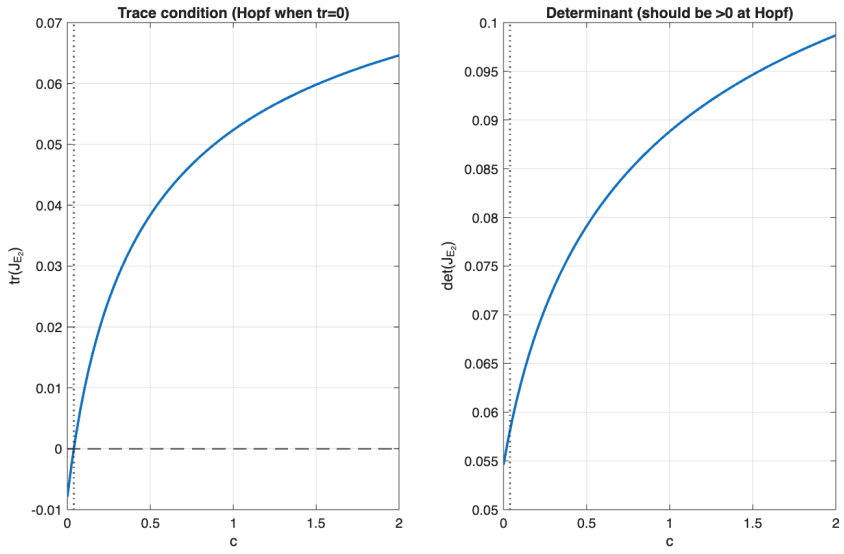


Figure 4: Numerical verification of the Hopf bifurcation conditions at the coexistence equilibrium E_2 with respect to the predator cooperation parameter c.

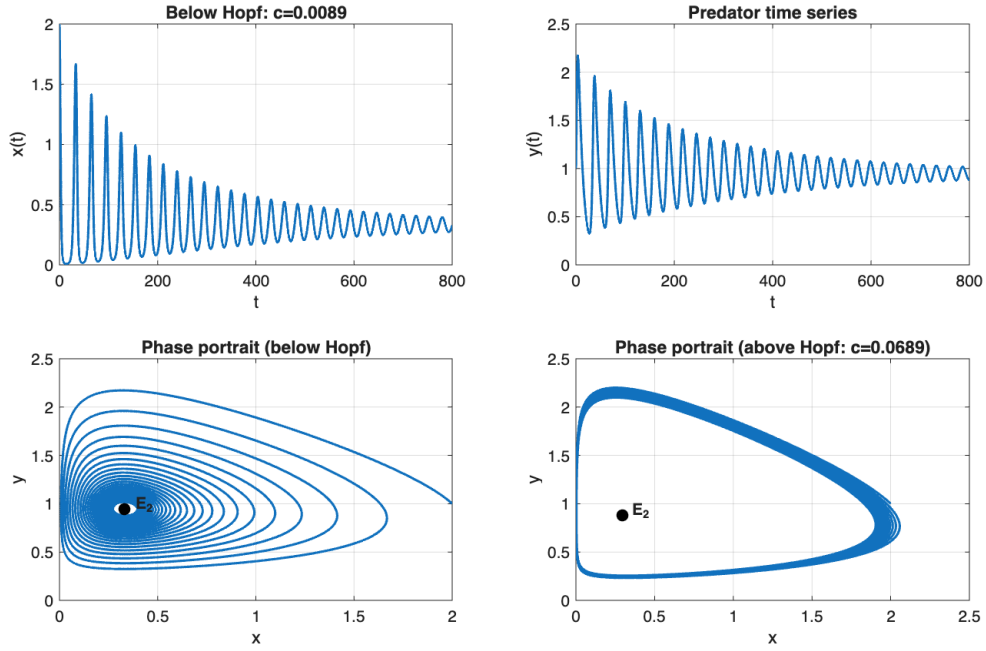


Figure 5: Numerical dynamics below and above the Hopf bifurcation threshold. Panel (a) shows the prey time series $x(t)$ for $c = 0.0089 < c^*$. Panel (b) shows the predator time series $y(t)$ for the same subcritical value $c = 0.0089 < c^*$. Panel (c) shows the corresponding phase portrait for $c = 0.0089 < c^*$, where trajectories spiral inward toward the coexistence equilibrium E_2, indicating local asymptotic stability. Panel (d) shows the phase portrait for $c = 0.0689 > c^*$, where trajectories approach a stable closed orbit, indicating the emergence of a stable limit cycle after Hopf bifurcation.

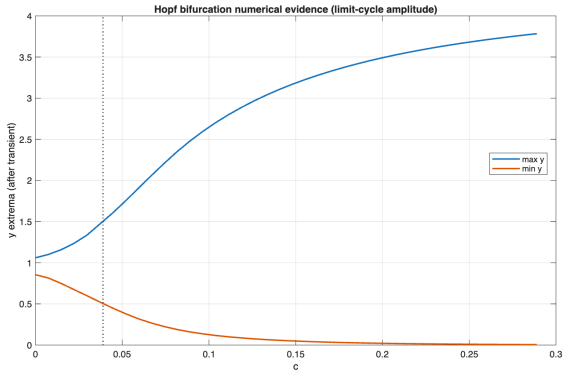


Figure 6: Numerical amplitude diagram illustrating oscillatory behavior as the predator cooperation parameter c varies.

5.1 Sensitivity of the Hopf threshold

To assess the robustness of the Hopf bifurcation established in Theorem 8, we examined how the critical predator cooperation threshold c^* varies with several ecologically important parameters.

In particular, we studied the dependence of c^* on the fear intensity k, the prey vigilance level v, the predator mortality rate δ_1, and the predator conversion efficiency γ. In each case, all remaining parameters were fixed at the baseline values used in the Hopf-bifurcation simulations,

$$r = 1, \qquad K = 10, \qquad \delta = 0.10, \qquad \mu = 0.50, \qquad q = 0.50, \qquad h = 0.20,$$

with the fixed parameters taken as

$$v = 0.20, \qquad k = 0.50, \qquad \delta_1 = 0.10, \qquad \gamma = 0.70.$$

For each selected parameter value, the critical threshold c^* was computed numerically by solving the coexistence equilibrium conditions together with the Hopf condition

$$\dot{x} = 0, \qquad \dot{y} = 0, \qquad \mathrm{tr}(J(E_2)) = 0,$$

while verifying that

$$\det(J(E_2)) > 0.$$

Figure 7 shows that the Hopf threshold varies smoothly and systematically across the tested parameter ranges, confirming that the bifurcation is not confined to a narrow parameter set. In panel (a), the threshold c^* increases with the fear parameter k. Thus, stronger fear effects require a larger degree of predator cooperation before oscillations can emerge, indicating that fear acts as a stabilizing mechanism. Panel (b) shows a similar increase of c^* with the vigilance level v, meaning that stronger prey vigilance delays the onset of destabilization by reducing predation efficiency and suppressing predator growth.

Panel (c) further shows that c^* increases with predator mortality δ_1. Ecologically, this means that when predators experience stronger natural losses, cooperation must be sufficiently strong to offset those losses before the coexistence equilibrium can lose stability. By contrast, panel (d) shows that c^* decreases as the conversion efficiency γ increases. Hence, when predators convert consumed prey into biomass more efficiently, even relatively small increases in cooperative hunting are enough to destabilize the coexistence equilibrium and generate oscillatory dynamics.

Overall, the sensitivity analysis quantitatively supports the biological conclusions of the model: Fear, vigilance, and predator mortality act as stabilizing factors because they raise the cooperation threshold required for predator-driven oscillations, whereas stronger predator demographic efficiency lowers that threshold and promotes instability.

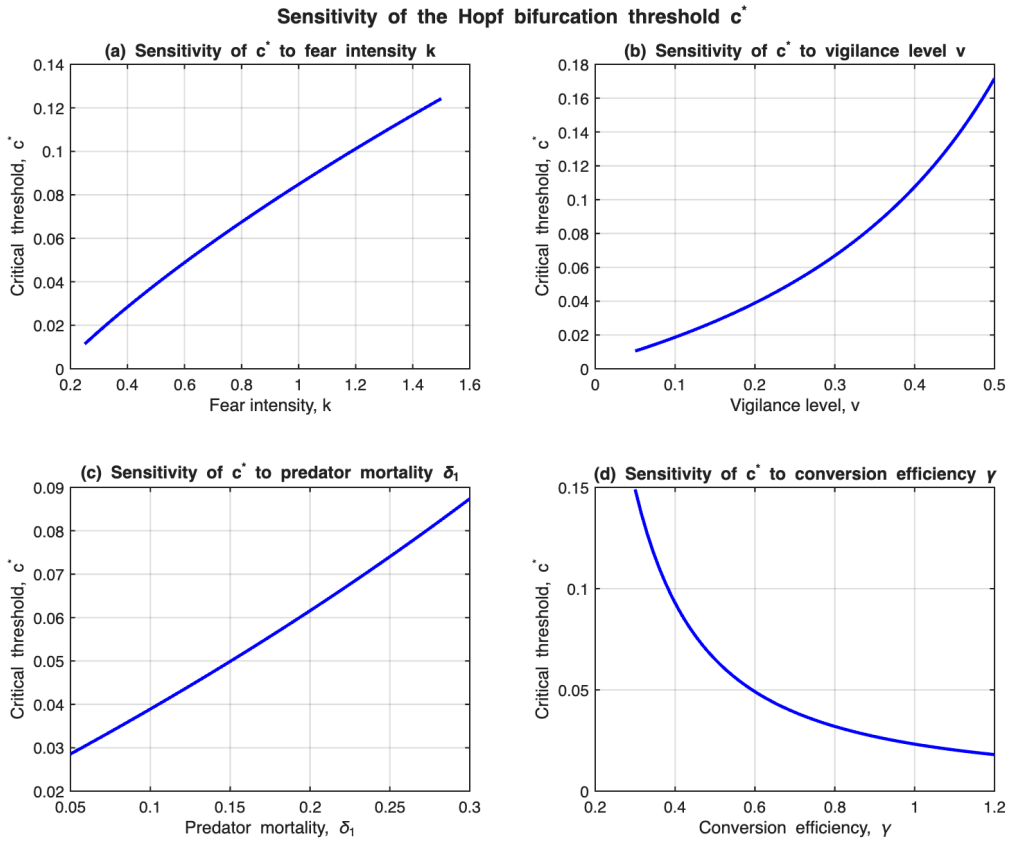


Figure 7: Sensitivity of the Hopf bifurcation threshold c^* to key parameters of system (1). Panel (a) shows c^* versus fear intensity k, panel (b) shows c^* versus vigilance level v, panel (c) shows c^* versus predator mortality δ_1, and panel (d) shows c^* versus conversion efficiency γ. In each panel, c^* is computed numerically from the coexistence conditions together with $\mathrm{tr}(J(E_2)) = 0$ and $\det(J(E_2)) > 0$, while all other parameters are fixed at the baseline values used in the Hopf-bifurcation simulations.

6 Conclusion

In this study, we proposed and rigorously analyzed a predator-prey model that integrates prey vigilance, fear effects, and predator cooperative hunting as key behavioral mechanisms governing population dynamics. Unlike classical predator-prey formulations, the present model explicitly accounts for the nonconsumptive effects of predation risk, whereby prey fear and vigilance reduce effective prey growth and alter interaction outcomes even in the absence of direct predation. This framework allows for a more realistic representation of predator-prey interactions observed in natural ecosystems.

We began by establishing the mathematical structure of the model and identifying all biologically meaningful equilibria, namely the extinction equilibrium, the predator-free equilibrium,

and the coexistence equilibrium. Local stability analysis revealed that prey fear and vigilance significantly modify the stability thresholds of these equilibria by effectively reducing prey reproductive potential and limiting predator access to prey. In particular, heightened fear responses were shown to suppress predator persistence by weakening the functional response, thereby enlarging the parameter regimes in which predator extinction occurs.

Beyond local analysis, we derived sufficient conditions for global asymptotic stability of both the extinction and predator-free equilibria using Lyapunov-based arguments. These results demonstrate that strong fear effects, in combination with prey vigilance or elevated predator mortality, can globally drive the system toward predator extinction regardless of initial population sizes. From an ecological standpoint, this highlights the powerful role of fear-mediated behavioral responses in shaping long-term population outcomes, independent of direct predation pressure.

The coexistence equilibrium was shown to exist under balanced levels of fear, vigilance, and cooperative hunting. Local stability analysis indicated that moderate predator cooperation stabilizes coexistence, while excessive cooperation may destabilize the equilibrium. This destabilization mechanism was formalized through the identification of a Hopf bifurcation with respect to the predator cooperation parameter. By analytically verifying the trace, determinant, and transversality conditions of the Jacobian matrix, we established that increasing cooperative hunting can induce a transition from stable coexistence to persistent oscillatory dynamics.

Extensive numerical simulations were conducted to support the theoretical findings. Phase-plane portraits confirmed the global stability of the extinction and predator-free equilibria under the derived conditions, while time-series simulations illustrated convergence to the coexistence equilibrium when locally stable. Near the Hopf threshold, numerical evidence clearly demonstrated the emergence of stable limit cycles, confirming the analytical bifurcation results. These oscillations reflect the combined effects of predator cooperation and prey fear, producing sustained fluctuations in population densities.

Overall, the results emphasize that fear effects are not merely secondary modifiers but fundamental drivers of predator–prey dynamics. When combined with prey vigilance and predator cooperation, fear can suppress predator persistence, stabilize equilibria, or generate complex oscillatory behavior depending on the parameter regime. The proposed model therefore provides a flexible and biologically grounded framework for investigating nonlethal predator effects and cooperative behaviors in ecological systems. A sensitivity analysis of the Hopf threshold further showed that increasing fear intensity, vigilance, and predator mortality shifts the critical cooperation level upward, whereas increasing predator conversion efficiency shifts it downward, thereby confirming that antipredator responses act as stabilizing mechanisms against predator-driven oscillations.

One limitation of the present formulation is that vigilance is assumed to affect prey demography primarily through reduced reproduction rather than through a direct increase in natural mortality. In some ecological systems, however, prolonged vigilance may indirectly elevate mortality through reduced body condition, physiological stress, immunosuppression, or increased accident risk during escape. Incorporating such mortality-mediated vigilance effects, together with extensions such as spatial diffusion, time delays, stochastic perturbations, or adaptive behavioral responses, would be a worthwhile direction for future research and could further enhance the applicability of the model to real-world ecological and conservation problems.

Funding

This research received no external funding.

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] Berec, L. *Impacts of foraging facilitation among predators on predator-prey dynamics*, Bull. Math. Biol. 72(1) (2010) 94–121.

[2] Boesch, C. *Cooperative hunting in wild chimpanzees*, Anim. Behav. 48(3) (1994) 653–667.

[3] Creel, S. and Christianson, D. *Relationships between direct predation and risk effects*, Trends Ecol. Evol. 23(4) (2008) 194–201.

[4] Creel, S., Christianson, D., Liley, S. and Winnie, J.A. *Predation risk affects reproductive physiology and demography of elk*, Science 315(5814) (2007) 960.

[5] Creel, S. and Creel, N.M. *Communal hunting and pack size in African wild dogs (*Lycaon pictus*)*, Anim. Behav. 50(5) (1995) 1325–1339.

[6] Delm, M.M. *Vigilance for predators: Detection and dilution effects*, Behav. Ecol. Sociobiol. 26(5) (1990) 337–342.

[7] Elgar, M.A. *Predator vigilance and group size in mammals and birds*, Biol. Rev. 64(1) (1989) 13–33.

[8] Fortin, D., Boyce, M.S., Merrill, E.H. and Fryxell, J.M. *Foraging costs of vigilance in large mammalian herbivores*, Oikos 107(1) (2004) 172–180.

[9] Illius, A.W. and FitzGibbon, C.D.F. *Costs of vigilance in foraging ungulates*, Anim. Behav. 47(2) (1994) 481–484.

[10] Lima, S.L. *Nonlethal effects in the ecology of predator-prey interactions*, BioScience 48(1) (1998) 25–34.

[11] Lima, S.L. and Dill, L.M. *Behavioral decisions made under the risk of predation*, Can. J. Zool. 68(4) (1990) 619–640.

[12] Lotka, A.J. *Elements of Physical Biology*, Williams & Wilkins, Baltimore, 1925.

[13] Packer, C. and Ruttan, L. *The evolution of cooperative hunting*, Am. Nat. 132(2) (1988) 159–198.

[14] Pal, S., Pal, N., Samanta, S. and Chattopadhyay, J. *Fear effect in prey and hunting cooperation among predators*, Math. Biosci. Eng. 16(5) (2019) 5146–5179.

[15] Peers, M.J.L., Majchrzak, Y.N., Neilson, E., Lamb, C.T., Hämäläinen, A., Haines, J.A. and Boutin, S. *Quantifying fear effects on prey demography in nature*, Ecology 99(8) (2018) 1716–1723.

[16] Périquet, S., Todd-Jones, L., Valeix, M., Stapelkamp, B., Elliot, N.B., Wijers, M. and Macdonald, D.W. *Influence of immediate predation risk by lions on the vigilance of prey of different body size*, Behav. Ecol. 23(5) (2012) 970–976.

[17] Scheel, D. and Packer, C. *Group hunting behaviour of lions: A search for cooperation*, Anim. Behav. 41(4) (1991) 697–709.

[18] Schmidt, P.A. and Mech, L.D. *Wolf pack size and food acquisition*, Am. Nat. 150(4) (1997) 513–517.

[19] Stander, P.E. *Cooperative hunting in lions: The role of the individual*, Behav. Ecol. Sociobiol. 29(6) (1992) 445–454.

[20] Suraci, J.P., Clinchy, M., Dill, L.M., Roberts, D. and Zanette, L.Y. *Fear of large carnivores causes a trophic cascade*, Nat. Commun. 7 (2016) 10698.

[21] Takyi, E.M., Ohanian, C., Cathcart, M. and Kumar, N. *Dynamical analysis of a predator-prey system with prey vigilance and hunting cooperation in predators*, Math. Biosci. Eng. 21(2) (2024) 2768–2786.

[22] Teixeira Alves, M.T. and Hilker, F.M. *Hunting cooperation and Allee effects in predators*, J. Theor. Biol. 419 (2017) 13–22.

[23] Volterra, V. *Fluctuations in the abundance of a species considered mathematically*, Nature 118 (1926) 558–560.

[24] Wang, X., Zanette, L.Y. and Zou, X. *Modelling the fear effect in predator-prey interactions*, J. Math. Biol. 73(5) (2016) 1179–1204.

[25] Watson, M., Aebischer, N.J. and Cresswell, W. *Vigilance and fitness in grey partridges*, J. Anim. Ecol. 76(1) (2007) 211–221.

Iranian Journal of Numerical Analysis and Optimization
Vol. 17, No. 1, 2027, pp 247–272
https://doi.org/10.22067/ijnao.2026.99197.1889
https://ijnao.um.ac.ir/

CC BY NC

Research Article

Algebraic properties of the new polyconvolution operator related to the Laplace, Fourier cosine integral transforms and applications

M.K. Nguyen*, L.T.P. Bui and V.T. Tran

Abstract

This work is focused on investigating a new polyconvolution operator with the trigonometric weight function $\beta(y_1) = \cos y_1$ associated with the Laplace and Fourier cosine integral transforms. We give an explicit form of this operator, prove the corresponding factorization identity, and establish some important algebraic properties of it, together with a Titchmarsh-type theorem. Within this analytic setting, we obtain fundamental properties including L_1-boundedness, Young-type inequalities on mixed L_p spaces, and a norm inequality of Saitoh's type in the $L_p(\mathbb{R}_+, q_i)$ weight spaces. Applications demonstrate the operator's effectiveness in

*Corresponding author

Received 6 June 2026; revised 29 August 2026; accepted 6 September 2026

Minh Khoa Nguyen
Department of Mathematics, Electric Power University, 235 Hoang Quoc Viet, Nghia Do, Hanoi, Viet Nam.
e-mail: khoanm@epu.edu.vn,

Le Trong Phuoc Bui
Department of Mathematics, Electric Power University, 235 Hoang Quoc Viet, Nghia Do, Hanoi, Viet Nam.
e-mail: phuocblt@epu.edu.vn,

Van Thang Tran
Department of Mathematics, Electric Power University, 235 Hoang Quoc Viet, Nghia Do, Hanoi, Viet Nam.
e-mail: thangtv@epu.edu.vn

How to cite this article
Nguyen, M.K., Bui, L.T.P. and Tran, V.T., Algebraic properties of the new polyconvolution operator related to the Laplace, Fourier cosine integral transforms and applications. *Iran. J. Numer. Anal. Optim.*, 2027; 17(1): 247-272. https://doi.org/10.22067/ijnao.2026.99197.1889

solving some new classes of integral equations of Toeplitz plus Hankel-type and of polyconvolution type, a system of equations of polyconvolution type, and an integro-differential equation of Barbashin-type, yielding explicit closed-form solutions.

AMS subject classifications (2020): 44A35, 44A10, 42A38, 45E10

Keywords: Fourier cosine, Laplace integral transforms, Convolution, Generalized convolution, polyconvolution, Integral equation, Integro-differential Barbashin equation.

1 Introduction

Research on convolutions and generalized convolutions for integral transforms has long played a central role in both theoretical and applied mathematics and in technical fields [1, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 16, 15, 23, 25, 28, 32, 30, 31, 36, 37, 27]. In 1997, with the idea of enlarging the convolution operator, Kakichev [14] advanced a generalized definition related to the polyconvolution operator for $n+1$ arbitrary integral transforms $G, G_1, G_2, \ldots, G_n$ with the weight function $\beta(y_1)$ of functions $g_1, g_2, \ldots, g_n$ for which the factorization identity holds:

$$G\big[\overset{\beta}{*}(g_1, g_2, \ldots, g_n)\big](y_1) = \beta(y_1)(G_1 g_1)(y_1) \cdot (G_2 g_2)(y_1) \cdots (G_n g_n)(y_1).$$

According to this idea, the polyconvolution operators related to the Fourier, Fourier cosine, Fourier sine, Hartley, and other types of Fourier integral transforms have been presented [16, 18, 19, 20, 21, 22, 29]; see also the polyconvolution operator for the Fourier cosine and Laplace integral transforms [35]. The generalized convolutions generated by the Fourier and Laplace transforms have been studied in [30, 31]. Following these authors, we built and studied a new polyconvolution operator with the weight function $\beta(y_1) = e^{-y_1}$ for a triple of integral transforms, the Laplace and Fourier cosine integral transforms [17]. In this direction, we build and study a new polyconvolution operator with the trigonometric weight function related to the Fourier cosine and Laplace integral transforms. Taking into account the above factorization equality, the general definition related to the polyconvolution operator has the following form:

$$\overset{\beta}{*}(g_1, g_2, \ldots, g_n)(x_1) = G^{-1}\big[\beta(\cdot)(G_1 g_1)(\cdot)(G_2 g_2)(\cdot) \cdots (G_n g_n)\big](x_1),$$

with G^{-1} being the inverse operator of the operator G. It is difficult to get an explicit form of a polyconvolution operator when applied to concrete integral transforms. The essence of this work is to introduce several interesting algebraic properties, to prove a Young-type theorem, some

norm inequalities in $L_p(\mathbb{R}_+)$ and a norm inequality of Saitoh's type in the $L_p(\mathbb{R}_+, q_i)$ weight spaces related to this polyconvolution operator, and apply this operator to several nonstandard integral equations and systems of equations of polyconvolution type, obtaining solutions in closed form. For such integral equations, a representation of the solution in closed form is an interesting open problem [3, 23]. Finally, with the help of this operator in subsection 5.4, we study a class of integro-differential equations of Barbashin's type [2, 3]; this is a concrete model arising in mathematical physics, radiation propagation and mathematical biology.

2 Preliminaries

In this section, we recall several known convolution and generalized convolution operators. The convolution operator of functions f_1, h_1 in the space $L_1(\mathbb{R}_+)$ associated with the Laplace operator is introduced in [25]:

$$\Big(f_1 \underset{L}{*} h_1\Big)(x_1) = \frac{1}{\sqrt{2\pi}} \int_0^{x_1} f_1(x_1 - y_1)\, h_1(y_1)\, dy_1, \quad x_1 > 0, \tag{1}$$

for which the following identity is satisfied:

$$L\Big(f_1 \underset{L}{*} h_1\Big)(y_1) = (Lf_1)(y_1)(Lh_1)(y_1), \quad y_1 > 0. \tag{2}$$

The symbol L denotes the Laplace integral operator [25]

$$(Lf_1)(y_1) = \int_0^{\infty} f_1(x_1)\, e^{-y_1 x_1}\, dx_1, \quad y_1 > 0.$$

The convolution operator of functions f_1 and h_1 in the space $L_1(\mathbb{R}_+)$ related to the Fourier cosine operator is introduced in [25, 33]:

$$\Big(f_1 \underset{F_c}{*} h_1\Big)(x_1) = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} f_1(y_1)\, [h_1(x_1 + y_1) + h_1(|x_1 - y_1|)]\, dy_1, \quad x_1 > 0, \tag{3}$$

with the identity

$$F_c\Big(f_1 \underset{F_c}{*} h_1\Big)(y_1) = (F_c f_1)(y_1)(F_c h_1)(y_1), \quad y_1 > 0, \tag{4}$$

where the Fourier cosine operator is defined by [25, 33]

$$(F_c f_1)(y_1) = \sqrt{\frac{2}{\pi}} \int_0^{\infty} f_1(x_1) \cos(x_1 y_1)\, dx_1, \quad y_1 > 0.$$

3 The polyconvolution operator with the trigonometric weight function related to the Fourier cosine and Laplace operators

In this section, we build the new polyconvolution operator with the trigonometric weight function $\beta(y_1) = \cos y_1$ related to the Fourier cosine and Laplace operators. We also study some operational properties of this operator.

Definition 1. The polyconvolution operator of functions f_1, g_1, h_1 with the weight function $\beta(y_1) = \cos y_1$ associated with the Fourier cosine and Laplace operators is denoted by $\overset{\beta}{\underset{F_c,L}{*}}(f_1, g_1, h_1)$ and defined by

$$\begin{aligned}
&\overset{\beta}{\underset{F_c,L}{*}}(f_1, g_1, h_1)(x_1) \\
&= \frac{1}{2\pi}\int_0^\infty\int_0^\infty\int_0^\infty \Bigg[\frac{v_1+w_1}{(v_1+w_1)^2+(u_1+x_1+1)^2} \\
&\quad + \frac{v_1+w_1}{(v_1+w_1)^2+(u_1-x_1+1)^2} + \frac{v_1+w_1}{(v_1+w_1)^2+(u_1+x_1-1)^2} \\
&\quad + \frac{v_1+w_1}{(v_1+w_1)^2+(u_1-x_1-1)^2}\Bigg] f_1(u_1)\, g_1(v_1)\, h_1(w_1)\, du_1\, dv_1\, dw_1, \quad x_1 > 0.
\end{aligned} \tag{5}$$

Theorem 1. Assume that f_1, g_1, h_1 are arbitrary functions in the space $L_1(\mathbb{R}_+)$. Then the operator (5) belongs to the space $L_1(\mathbb{R}_+)$ and satisfies the following property:

$$F_c\big[\overset{\beta}{\underset{F_c,L}{*}}(f_1, g_1, h_1)\big](y_1) = \cos y_1\,(F_c f_1)(y_1)(Lg_1)(y_1)(Lh_1)(y_1), \tag{6}$$

for all $y_1 > 0$.

Furthermore, $F_c\big[\overset{\beta}{\underset{F_c,L}{*}}(f_1, g_1, h_1)\big] \in C_0(\mathbb{R}_+)$; here $C_0(\mathbb{R}_+)$ is the space of bounded continuous functions vanishing at infinity. On the other hand, the L_1-norm inequality of this polyconvolution operator is as follows:

$$\big\|\overset{\beta}{\underset{F_c,L}{*}}(f_1, g_1, h_1)\big\|_{L_1(\mathbb{R}_+)} \le \|f_1\|_{L_1(\mathbb{R}_+)} \cdot \|g_1\|_{L_1(\mathbb{R}_+)} \cdot \|h_1\|_{L_1(\mathbb{R}_+)}. \tag{7}$$

Finally, when $f_1, g_1, h_1 \in L_2(\mathbb{R}_+) \cap L_1(\mathbb{R}_+)$, the L_2-inversion formula holds, the equality being understood in the $L_2(\mathbb{R}_+)$-sense:

$$\overset{\beta}{\underset{F_c,L}{*}}(f_1, g_1, h_1)(x_1) \tag{8}$$

$$= \sqrt{\frac{2}{\pi}} \int_0^\infty \cos y_1 \, (F_c f_1)(y_1)(Lg_1)(y_1)(Lh_1)(y_1) \cos(x_1 y_1) \, dy_1.$$

Proof. Using the formula (5), the Tonelli theorem, and the Fubini theorem, we have

$$\begin{aligned}
&\big\| \overset{\beta}{*}_{F_c,L} (f_1, g_1, h_1) \big\|_{L_1(\mathbb{R}_+)} \\
&= \int_0^\infty \big| \overset{\beta}{*}_{F_c,L} (f_1, g_1, h_1)(x_1) \big| \, dx_1 \\
&\le \frac{1}{2\pi} \int_0^\infty \int_0^\infty \int_0^\infty \int_0^\infty \left| \frac{v_1 + w_1}{(v_1 + w_1)^2 + (u_1 + x_1 + 1)^2} \right. \\
&\quad + \frac{v_1 + w_1}{(v_1 + w_1)^2 + (u_1 - x_1 + 1)^2} + \frac{v_1 + w_1}{(v_1 + w_1)^2 + (u_1 + x_1 - 1)^2} \\
&\quad \left. + \frac{v_1 + w_1}{(v_1 + w_1)^2 + (u_1 - x_1 - 1)^2} \right| \\
&\quad \times |f_1(u_1)| \cdot |g_1(v_1)| \cdot |h_1(w_1)| \, du_1 \, dv_1 \, dw_1 \, dx_1 \\
&= \frac{1}{2\pi} \int_0^\infty \left| \frac{v_1 + w_1}{(v_1 + w_1)^2 + (u_1 + x_1 + 1)^2} + \frac{v_1 + w_1}{(v_1 + w_1)^2 + (u_1 - x_1 + 1)^2} \right. \\
&\quad + \frac{v_1 + w_1}{(v_1 + w_1)^2 + (u_1 + x_1 - 1)^2} + \left. \frac{v_1 + w_1}{(v_1 + w_1)^2 + (u_1 - x_1 - 1)^2} \right| dx_1 \\
&\quad \cdot \int_0^\infty |f_1(u_1)| \, du_1 \cdot \int_0^\infty |g_1(v_1)| \, dv_1 \cdot \int_0^\infty |h_1(w_1)| \, dw_1.
\end{aligned}$$

We consider

$$\begin{aligned}
I_1 = \int_0^\infty \left[\frac{v_1 + w_1}{(v_1 + w_1)^2 + (u_1 + x_1 + 1)^2} + \frac{v_1 + w_1}{(v_1 + w_1)^2 + (u_1 - x_1 + 1)^2} \right. \\
\left. + \frac{v_1 + w_1}{(v_1 + w_1)^2 + (u_1 + x_1 - 1)^2} + \frac{v_1 + w_1}{(v_1 + w_1)^2 + (u_1 - x_1 - 1)^2} \right] dx_1.
\end{aligned}$$

Then I_1 can be obtained by pairing the two translated Poisson-type kernels. Let $a = v_1 + w_1 > 0$. Then

$$\int_0^\infty \left[\frac{a}{a^2 + (u_1 + x_1 + 1)^2} + \frac{a}{a^2 + (u_1 - x_1 + 1)^2} \right] dx_1 = \pi,$$

and the analogous pair involving $u_1 \pm x_1 - 1$ also has integral π. Hence the integral I_1 of the full nonnegative kernel with respect to x_1 equals 2π.

So,

$$\big\| \overset{\beta}{*}_{F_c,L} (f_1, g_1, h_1) \big\|_{L_1(\mathbb{R}_+)} \le \|f_1\|_{L_1(\mathbb{R}_+)} \cdot \|g_1\|_{L_1(\mathbb{R}_+)} \cdot \|h_1\|_{L_1(\mathbb{R}_+)} < +\infty.$$

Therefore, we have (7) and $\overset{\beta}{*}_{F_c,L} (f_1, g_1, h_1) \in L_1(\mathbb{R}_+)$.

Alternatively, using (5) and the formula $\int_0^\infty e^{-ax_1}\cos y_1x_1\,dx_1 = \dfrac{a}{a^2+y_1^2}$, $a>0$ (see [11, formula 2.13.5]), we get

$$\begin{aligned}
&\sqrt{\frac{2}{\pi}}\int_0^\infty \big[\cos y_1\,(F_cf_1)(y_1)(Lg_1)(y_1)(Lh_1)(y_1)\big]\cos y_1x_1\,dy_1\\
&=\sqrt{\frac{2}{\pi}}\int_0^\infty\Bigg[\cos y_1\sqrt{\frac{2}{\pi}}\int_0^\infty f_1(u_1)\cos y_1u_1\,du_1\int_0^\infty g_1(v_1)e^{-y_1v_1}\,dv_1\\
&\quad\times\int_0^\infty h_1(w_1)e^{-y_1w_1}\,dw_1\Bigg]\cos y_1x_1\,dy_1\\
&=\frac{2}{\pi}\int_0^\infty\int_0^\infty\int_0^\infty\int_0^\infty\cos y_1\cos y_1u_1\cos y_1x_1\cdot e^{-y_1(v_1+w_1)}\\
&\quad\times f_1(u_1)\,g_1(v_1)\,h_1(w_1)\,du_1\,dv_1\,dw_1\,dy_1\\
&=\frac{1}{2\pi}\int_0^\infty\int_0^\infty\int_0^\infty\int_0^\infty\big[\cos y_1(u_1+x_1+1)+\cos y_1(u_1+x_1-1)\\
&\quad+\cos y_1(u_1-x_1+1)+\cos y_1(u_1-x_1-1)\big]e^{-y_1(v_1+w_1)}\\
&\quad\times f_1(u_1)\cdot g_1(v_1)\cdot h_1(w_1)\,du_1\,dv_1\,dw_1\,dy_1\\
&=\overset{\beta}{\underset{F_c,L}{*}}(f_1,g_1,h_1)(x_1).
\end{aligned}$$

Thus, when $f_1,g_1,h_1\in L_2(\mathbb{R}_+)\cap L_1(\mathbb{R}_+)$ the L_2-inversion formula (8) holds. Combining this with $\overset{\beta}{\underset{F_c,L}{*}}(f_1,g_1,h_1)\in L_1(\mathbb{R}_+)$, applying the Fourier cosine integral transform to both sides of (8), we obtain the factorization identity (6).

The Riemann–Lebesgue theorem in [26] states that if $f_1\in L_1(\mathbb{R}^n)$, then $(Ff_1)(y_1)\to 0$ as $|y_1|\to+\infty$, and hence $(Ff_1)\in C_0(\mathbb{R}^n)$. This is still true for the Fourier cosine integral transform F_c [33, 11]. So, $F_c\big[\overset{\beta}{\underset{F_c,L}{*}}(f_1,g_1,h_1)\big]$ belongs to $C_0(\mathbb{R}_+)$ by virtue of the Riemann–Lebesgue theorem. □

Proposition 1. Assume that $f_1,g_1,l_1,k_1\in L_1(\mathbb{R}_+)$. Then, the polyconvolution (5) satisfies the following equalities:

a) $\overset{\beta}{\underset{F_c,L}{*}}(f_1,g_1,l_1)=\overset{\beta}{\underset{F_c,L}{*}}(f_1,l_1,g_1)$.

Furthermore, we denote by D the class of functions on $\mathbb{R}_+$ that are piecewise continuous and satisfy an estimate: $|f_1(t)|\le Ce^{\alpha t}$ for some $C>0$, $\alpha\in\mathbb{R}$. Hence, if $f_1,g_1,l_1,k_1\in L_1(\mathbb{R}_+)\cap D$, we have the following equations:

b) $\overset{\beta}{\underset{F_c,L}{*}}\big[f_1,(g_1\underset{L}{*}l_1),k_1\big]=\overset{\beta}{\underset{F_c,L}{*}}\big[f_1,g_1,(l_1\underset{L}{*}k_1)\big]$,

c) $\overset{\beta}{\underset{F_c,L}{*}}\big[f_1,(g_1\underset{L}{*}l_1),k_1\big]=\overset{\beta}{\underset{F_c,L}{*}}\big[f_1,k_1,(g_1\underset{L}{*}l_1)\big]$.

Proof. First, we prove a). Indeed, from the factorization equality (6), we obtain

$$\begin{aligned} F_c[\overset{\beta}{\underset{F_c,L}{*}}(f_1,g_1,l_1)](y_1) &= \cos y_1\,(F_cf_1)(y_1)(Lg_1)(y_1)(Ll_1)(y_1)\\ &= \cos y_1\,(F_cf_1)(y_1)(Ll_1)(y_1)(Lg_1)(y_1)\\ &= F_c[\overset{\beta}{\underset{F_c,L}{*}}(f_1,l_1,g_1)](y_1). \end{aligned}$$

Thus $\overset{\beta}{\underset{F_c,L}{*}}(f_1,g_1,l_1) = \overset{\beta}{\underset{F_c,L}{*}}(f_1,l_1,g_1)$.

b) Using the factorization properties (2), (6), and $g_1, l_1 \in D \cap L_1(\mathbb{R}_+)$, we can write

$$\begin{aligned} F_c\big\{\overset{\beta}{\underset{F_c,L}{*}}[f_1,(g_1\underset{L}{*}l_1),k_1]\big\}(y_1) &= \cos y_1\,(F_cf_1)(y_1)\,L(g_1\underset{L}{*}l_1)(y_1)(Lk_1)(y_1)\\ &= \cos y_1\,(F_cf_1)(y_1)(Lg_1)(y_1)(Ll_1)(y_1)(Lk_1)(y_1)\\ &= \cos y_1\,(F_cf_1)(y_1)(Lg_1)(y_1)\,L(l_1\underset{L}{*}k_1)(y_1)\\ &= F_c\big\{\overset{\beta}{\underset{F_c,L}{*}}[f_1,g_1,(l_1\underset{L}{*}k_1)]\big\}(y_1), \quad y_1>0. \end{aligned}$$

Thus, $\overset{\beta}{\underset{F_c,L}{*}}\big[f_1,(g_1\underset{L}{*}l_1),k_1\big] = \overset{\beta}{\underset{F_c,L}{*}}\big[f_1,g_1,(l_1\underset{L}{*}k_1)\big]$. Similarly, we can prove c). □

Theorem 2 (Titchmarsh-type theorem)**.** Suppose $f_1 \in L_1(\mathbb{R}_+, e^{\alpha x_1})$, $\alpha > 0$ and $g_1, h_1 \in L_1(\mathbb{R}_+)$. If $\overset{\beta}{\underset{F_c,L}{*}}(f_1,g_1,h_1)(x_1) = 0$, for all $x_1 > 0$, then either $f_1(x_1) = 0$, for all $x_1 > 0$ or $g_1(x_1) = 0$, for all $x_1 > 0$ or $h_1(x_1) = 0$, for all $x_1 > 0$.

Proof. Indeed, from the hypothesis $\overset{\beta}{\underset{F_c,L}{*}}(f_1,g_1,h_1)(x_1) = 0$, for all $x_1 > 0$ and the factorization identity (6), we have

$$\cos y_1\,(F_cf_1)(y_1)(Lg_1)(y_1)(Lh_1)(y_1) = 0, \quad \text{for all } y_1 > 0. \tag{9}$$

On the other hand, we have

$$\begin{aligned} \left|\frac{d^n}{dy_1^n}\big[\cos(x_1y_1)\,f_1(x_1)\big]\right| &= \left|f_1(x_1)\cdot x_1^n\cdot\cos\left(x_1y_1+n\frac{\pi}{2}\right)\right|\\ &\le \left|e^{-\alpha x_1}\cdot x_1^n\right|\cdot\left|e^{\alpha x_1}\cdot f_1(x_1)\right| \le \frac{n!}{\alpha^n}\cdot\left|e^{\alpha x_1}\cdot f_1(x_1)\right|. \end{aligned}$$

Here, we use the estimation

$$0 \le e^{-\alpha x_1}\cdot x_1^n = e^{-\alpha x_1}\cdot\frac{(\alpha x_1)^n}{n!}\cdot\frac{n!}{\alpha^n} \le e^{-\alpha x_1}\cdot e^{\alpha x_1}\cdot\frac{n!}{\alpha^n} = \frac{n!}{\alpha^n}.$$

Since $f_1 \in L_1(\mathbb{R}_+, e^{\alpha x_1}) \subset L_1(\mathbb{R}_+)$, we conclude that $(F_c f_1)$ is analytic in $\mathbb{R}_+$. Moreover, $(Lg_1)(y_1)$, $(Lh_1)(y_1)$ are analytic in $\mathbb{R}_+$. Together with (9), it implies that either $f_1(x_1) = 0$, for all $x_1 > 0$, or $g_1(x_1) = 0$, for all $x_1 > 0$, or $h_1(x_1) = 0$, for all $x_1 > 0$. □

4 Inequalities related to the polyconvolution operator

4.1 Norm inequality in $L_p(\mathbb{R}_+, e^{-\alpha x_1})$

The standard norms are defined as follows:

$$\|f_1\|_{L_1(\mathbb{R}_+)} = \int_0^\infty |f_1(x_1)|\, dx_1;$$
$$\|f_1\|_{L_1(\mathbb{R}_+,q)} = \int_0^\infty |f_1(x_1)| q(x_1)\, dx_1, \qquad q(x_1) > 0, \quad \text{for all } x_1 > 0;$$
$$\|f_1\|_{L_p(\mathbb{R}_+)} = \left[\int_0^\infty |f_1(x_1)|^p\, dx_1\right]^{1/p};$$
$$\|f_1\|_{L_p(\mathbb{R}_+,q)} = \left[\int_0^\infty |f_1(x_1)|^p q(x_1)\, dx_1\right]^{1/p}, \qquad q(x_1) > 0, \quad \text{for all } x_1 > 0.$$

In this subsection, we establish the norm inequalities for the polyconvolution (5) in $L_p(\mathbb{R}_+, e^{-\alpha x_1})$ with $1 \le p \le \infty$.

Theorem 3. Let $f_1 \in L_p(\mathbb{R}_+)$, $g_1 \in L_q(\mathbb{R}_+)$, $h_1 \in L_r(\mathbb{R}_+)$ be such that $p, q, r > 1$ and $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 2$. Then the polyconvolution (5) is bounded in $L_s(\mathbb{R}_+, e^{-\alpha x_1})$ when $s > 1$, $\alpha > 0$ and the following inequality holds:

$$\left\| \overset{\beta}{\underset{F_c,L}{*}} (f_1, g_1, h_1) \right\|_{L_s(\mathbb{R}_+, e^{-\alpha x_1})} \le \frac{1}{\alpha^{1/s}} \|f_1\|_{L_p(\mathbb{R}_+)} \|g_1\|_{L_q(\mathbb{R}_+)} \|h_1\|_{L_r(\mathbb{R}_+)}. \tag{10}$$

Proof. From the proof of Theorem 1, we have the following inequality:

$$\left| \overset{\beta}{\underset{F_c,L}{*}} (f_1, g_1, h_1)(x_1) \right| \le \int_0^\infty \int_0^\infty \int_0^\infty |f_1(u_1)||g_1(v_1)||h_1(w_1)|\, du_1\, dv_1\, dw_1, \qquad x_1 \in \mathbb{R}_+.$$

Let p_1, q_1, r_1 be the conjugate exponents of p, q, r and let

$$M = |g_1(v_1)|^{\frac{q}{p_1}} \cdot |h_1(w_1)|^{\frac{r}{p_1}},$$
$$N = |h_1(w_1)|^{\frac{r}{q_1}} \cdot |f_1(u_1)|^{\frac{p}{q_1}},$$
$$P = |f_1(u_1)|^{\frac{p}{r_1}} \cdot |g_1(v_1)|^{\frac{q}{r_1}}.$$

We see that $MNP = |f_1(u_1)| \cdot |g_1(v_1)| \cdot |h_1(w_1)|$.

Using Fubini's theorem, we get

$$\begin{aligned}\|M\|^{p_1}_{L_{p_1}(\mathbb{R}^3_+)} &= \int_0^\infty \int_0^\infty \left\{|g_1(v_1)|^{q/p_1} |h_1(w_1)|^{r/p_1}\right\}^{p_1} dv_1\, dw_1 \\ &= \int_0^\infty |g_1(v_1)|^q \Big(\int_0^\infty |h_1(w_1)|^r\, dw_1 \Big) dv_1 \\ &= \int_0^\infty |g_1(v_1)|^q \, \|h_1\|^r_{L_r(\mathbb{R}_+)}\, dv_1 = \|g_1\|^q_{L_q(\mathbb{R}_+)} \|h_1\|^r_{L_r(\mathbb{R}_+)}.\end{aligned}$$

Similarly, we get

$$\|N\|^{q_1}_{L_{q_1}(\mathbb{R}^3_+)} = \|f_1\|^p_{L_p(\mathbb{R}_+)} \|h_1\|^r_{L_r(\mathbb{R}_+)}; \qquad \|P\|^{r_1}_{L_{r_1}(\mathbb{R}^3_+)} = \|f_1\|^p_{L_p(\mathbb{R}_+)} \|g_1\|^q_{L_q(\mathbb{R}_+)}. \tag{11}$$

Since $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 2$, we have $\frac{1}{p_1} + \frac{1}{q_1} + \frac{1}{r_1} = 1$. Using Hölder's inequality and (11), we obtain

$$\begin{aligned}\Big| \overset{\beta}{\underset{F_c,L}{*}} (f_1, g_1, h_1)\Big| &\le \int_0^\infty \int_0^\infty \int_0^\infty MNP\, du_1\, dv_1\, dw_1 \\ &\le \left(\int_0^\infty \int_0^\infty \int_0^\infty M^{p_1}\, du_1\, dv_1\, dw_1 \right)^{1/p_1} \\ &\quad \times \left(\int_0^\infty \int_0^\infty \int_0^\infty N^{q_1}\, du_1\, dv_1\, dw_1 \right)^{1/q_1} \\ &\quad \times \left(\int_0^\infty \int_0^\infty \int_0^\infty P^{r_1}\, du_1\, dv_1\, dw_1 \right)^{1/r_1} \\ &= \|M\|_{L_{p_1}(\mathbb{R}^3_+)} \|N\|_{L_{q_1}(\mathbb{R}^3_+)} \|P\|_{L_{r_1}(\mathbb{R}^3_+)} \\ &= \left(\|g_1\|^{q/p_1}_{L_q(\mathbb{R}_+)} \|h_1\|^{r/p_1}_{L_r(\mathbb{R}_+)} \right) \left(\|f_1\|^{p/q_1}_{L_p(\mathbb{R}_+)} \|h_1\|^{r/q_1}_{L_r(\mathbb{R}_+)} \right) \\ &\quad \times \left(\|f_1\|^{p/r_1}_{L_p(\mathbb{R}_+)} \|g_1\|^{q/r_1}_{L_q(\mathbb{R}_+)} \right) \\ &= \|f_1\|_{L_p(\mathbb{R}_+)} \|g_1\|_{L_q(\mathbb{R}_+)} \|h_1\|_{L_r(\mathbb{R}_+)}.\end{aligned}$$

Since $\int_0^\infty e^{-\alpha x_1}\, dx_1 = \frac{1}{\alpha}$, $\alpha > 0$, we have

$$\begin{aligned}&\Big\| \overset{\beta}{\underset{F_c,L}{*}} (f_1, g_1, h_1)\Big\|^s_{L_s(\mathbb{R}_+, e^{-\alpha x_1})} \\ &= \int_0^\infty e^{-\alpha x_1} \Big| \overset{\beta}{\underset{F_c,L}{*}} (f_1, g_1, h_1)(x_1)\Big|^s dx_1\end{aligned}$$

$$\leq \int_0^\infty e^{-\alpha x_1} \|f_1\|^s_{L_p(\mathbb{R}_+)} \|g_1\|^s_{L_q(\mathbb{R}_+)} \|h_1\|^s_{L_r(\mathbb{R}_+)} \, dx_1$$
$$= \frac{1}{\alpha} \|f_1\|^s_{L_p(\mathbb{R}_+)} \|g_1\|^s_{L_q(\mathbb{R}_+)} \|h_1\|^s_{L_r(\mathbb{R}_+)}.$$

So, we obtain

$$\Big\| \overset{\beta}{\underset{F_c,L}{*}} (f_1, g_1, h_1) \Big\|_{L_s(\mathbb{R}_+, e^{-\alpha x_1})} \leq \frac{1}{\alpha^{1/s}} \|f_1\|_{L_p(\mathbb{R}_+)} \|g_1\|_{L_q(\mathbb{R}_+)} \|h_1\|_{L_r(\mathbb{R}_+)}.$$

□

4.2 The Young-type theorem related to this operator

Theorem 4. Suppose that $p, q, r, m > 1$ satisfy the condition $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} + \frac{1}{m} = 3$. If $f_1 \in L_p(\mathbb{R}_+)$, $g_1 \in L_q(\mathbb{R}_+)$, $h_1 \in L_r(\mathbb{R}_+)$, $k_1 \in L_m(\mathbb{R}_+)$, then the following estimate holds:

$$\begin{aligned} &\left| \int_0^\infty \Big[\overset{\beta}{\underset{F_c,L}{*}} (f_1, g_1, h_1)(x_1) \cdot k_1(x_1) \Big] \, dx_1 \right| \\ &\quad \leq \|f_1\|_{L_p(\mathbb{R}_+)} \cdot \|g_1\|_{L_q(\mathbb{R}_+)} \cdot \|h_1\|_{L_r(\mathbb{R}_+)} \cdot \|k_1\|_{L_m(\mathbb{R}_+)}. \end{aligned} \tag{12}$$

Proof. From (5) and the proof of Theorem 1, we have the estimate

$$\begin{aligned} &\left| \int_0^\infty \Big[\overset{\beta}{\underset{F_c,L}{*}} (f_1, g_1, h_1)(x_1) \cdot k_1(x_1) \Big] \, dx_1 \right| \\ &\quad \leq \int_0^\infty \int_0^\infty \int_0^\infty \int_0^\infty |f_1(u_1)| \cdot |g_1(v_1)| \cdot |h_1(w_1)| \cdot |k_1(x_1)| dw_1 \, dv_1 \, du_1 \, dx_1. \end{aligned} \tag{13}$$

Let p_1, q_1, r_1, m_1 be the conjugate exponents of p, q, r, m, respectively. Then

$$\frac{1}{p} + \frac{1}{p_1} = \frac{1}{q} + \frac{1}{q_1} = \frac{1}{r} + \frac{1}{r_1} = \frac{1}{m} + \frac{1}{m_1} = 1.$$

From the condition $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} + \frac{1}{m} = 3$, we obtain that

$$\begin{cases} \dfrac{1}{p_1} + \dfrac{1}{q_1} + \dfrac{1}{r_1} + \dfrac{1}{m_1} = 1, \\ p\Big(\dfrac{1}{q_1} + \dfrac{1}{r_1} + \dfrac{1}{m_1}\Big) = q\Big(\dfrac{1}{p_1} + \dfrac{1}{r_1} + \dfrac{1}{m_1}\Big) = r\Big(\dfrac{1}{p_1} + \dfrac{1}{q_1} + \dfrac{1}{m_1}\Big) \\ \qquad = m\Big(\dfrac{1}{p_1} + \dfrac{1}{q_1} + \dfrac{1}{r_1}\Big) = 1. \end{cases} \tag{14}$$

Putting

$$\begin{aligned}
K_1 &= |g_1(v_1)|^{\frac{q}{p_1}} \cdot |h_1(w_1)|^{\frac{r}{p_1}} \cdot |k_1(x_1)|^{\frac{m}{p_1}},\\
K_2 &= |f_1(u_1)|^{\frac{p}{q_1}} \cdot |h_1(w_1)|^{\frac{r}{q_1}} \cdot |k_1(x_1)|^{\frac{m}{q_1}},\\
K_3 &= |k_1(x_1)|^{\frac{m}{r_1}} \cdot |f_1(u_1)|^{\frac{p}{r_1}} \cdot |g_1(v_1)|^{\frac{q}{r_1}},\\
K_4 &= |f_1(u_1)|^{\frac{p}{m_1}} \cdot |g_1(v_1)|^{\frac{q}{m_1}} \cdot |h_1(w_1)|^{\frac{r}{m_1}}.
\end{aligned}$$

From (14), we obtain

$$K_1 \cdot K_2 \cdot K_3 \cdot K_4 = |f_1(u_1)| \cdot |g_1(v_1)| \cdot |h_1(w_1)| \cdot |k_1(x_1)|. \tag{15}$$

From (15), we have

$$\begin{aligned}
&\left| \int_0^\infty \left[{}^{\beta}_{*}_{F_c,L} (f_1, g_1, h_1)(x_1) \cdot k_1(x_1) \right] dx_1 \right| \\
&\quad \le \int_0^\infty \int_0^\infty \int_0^\infty \int_0^\infty K_1 \cdot K_2 \cdot K_3 \cdot K_4 \, dx_1 \, du_1 \, dv_1 \, dw_1.
\end{aligned}$$

Since $\frac{1}{p_1} + \frac{1}{q_1} + \frac{1}{r_1} + \frac{1}{m_1} = 1$, using Hölder's inequality, we have the following estimate:

$$\begin{aligned}
&\left| \int_0^\infty \left[{}^{\beta}_{*}_{F_c,L} (f_1, g_1, h_1)(x_1) \cdot k_1(x_1) \right] dx_1 \right| \\
&\le \left(\int_0^\infty \int_0^\infty \int_0^\infty \int_0^\infty |K_1|^{p_1} \, dx_1 \, du_1 \, dv_1 \, dw_1 \right)^{\frac{1}{p_1}} \\
&\quad \cdot \left(\int_0^\infty \int_0^\infty \int_0^\infty \int_0^\infty |K_2|^{q_1} \, dx_1 \, du_1 \, dv_1 \, dw_1 \right)^{\frac{1}{q_1}} \\
&\quad \cdot \left(\int_0^\infty \int_0^\infty \int_0^\infty \int_0^\infty |K_3|^{r_1} \, dx_1 \, du_1 \, dv_1 \, dw_1 \right)^{\frac{1}{r_1}} \\
&\quad \cdot \left(\int_0^\infty \int_0^\infty \int_0^\infty \int_0^\infty |K_4|^{m_1} \, dx_1 \, du_1 \, dv_1 \, dw_1 \right)^{\frac{1}{m_1}} \\
&= \|K_1\|_{L_{p_1}(\mathbb{R}_+^4)} \cdot \|K_2\|_{L_{q_1}(\mathbb{R}_+^4)} \cdot \|K_3\|_{L_{r_1}(\mathbb{R}_+^4)} \cdot \|K_4\|_{L_{m_1}(\mathbb{R}_+^4)}.
\end{aligned} \tag{16}$$

Furthermore, for $g_1 \in L_q(\mathbb{R}_+)$, $h_1 \in L_r(\mathbb{R}_+)$, $k_1 \in L_m(\mathbb{R}_+)$, along with the help of Fubini theorem, we have

$$\begin{aligned}\|K_1\|^{p_1}_{L_{p_1}(\mathbb{R}^4_+)} &= \int_0^\infty\int_0^\infty\int_0^\infty |g_1(v_1)|^q\cdot|h_1(w_1)|^r\cdot|k_1(x_1)|^m\,dv_1\,dw_1\,dx_1\\ &\le \int_0^\infty |g_1(v_1)|^q\,dv_1\cdot\int_0^\infty |h_1(w_1)|^r\,dw_1\cdot\int_0^\infty |k_1(x_1)|^m\,dx_1\\ &= \|g_1\|^q_{L_q(\mathbb{R}_+)}\cdot\|h_1\|^r_{L_r(\mathbb{R}_+)}\cdot\|k_1\|^m_{L_m(\mathbb{R}_+)}.\end{aligned}$$

So, we have the norm estimate K_1 as follows:

$$\|K_1\|_{L_{p_1}(\mathbb{R}^4_+)} \le \|g_1\|^{\frac{q}{p_1}}_{L_q(\mathbb{R}_+)}\cdot\|h_1\|^{\frac{r}{p_1}}_{L_r(\mathbb{R}_+)}\cdot\|k_1\|^{\frac{m}{p_1}}_{L_m(\mathbb{R}_+)}. \tag{17}$$

Similarly, we get

$$\|K_2\|_{L_{q_1}(\mathbb{R}^4_+)} \le \|h_1\|^{\frac{r}{q_1}}_{L_r(\mathbb{R}_+)}\cdot\|k_1\|^{\frac{m}{q_1}}_{L_m(\mathbb{R}_+)}\cdot\|f_1\|^{\frac{p}{q_1}}_{L_p(\mathbb{R}_+)}, \tag{18}$$

$$\|K_3\|_{L_{r_1}(\mathbb{R}^4_+)} \le \|k_1\|^{\frac{m}{r_1}}_{L_m(\mathbb{R}_+)}\cdot\|f_1\|^{\frac{p}{r_1}}_{L_p(\mathbb{R}_+)}\cdot\|g_1\|^{\frac{q}{r_1}}_{L_q(\mathbb{R}_+)}, \tag{19}$$

$$\|K_4\|_{L_{m_1}(\mathbb{R}^4_+)} \le \|f_1\|^{\frac{p}{m_1}}_{L_p(\mathbb{R}_+)}\cdot\|g_1\|^{\frac{q}{m_1}}_{L_q(\mathbb{R}_+)}\cdot\|h_1\|^{\frac{r}{m_1}}_{L_r(\mathbb{R}_+)}. \tag{20}$$

From (16)–(20), we obtain

$$\begin{aligned}&\left|\int_0^\infty \left[\overset{\beta}{\underset{F_c,L}{*}}(f_1,g_1,h_1)(x_1)\cdot k_1(x_1)\right]dx_1\right|\\ &\quad\le \|f_1\|_{L_p(\mathbb{R}_+)}\cdot\|g_1\|_{L_q(\mathbb{R}_+)}\cdot\|h_1\|_{L_r(\mathbb{R}_+)}\cdot\|k_1\|_{L_m(\mathbb{R}_+)}.\end{aligned}$$

□

4.3 Norm inequality of Saitoh's type in the $L_p(\mathbb{R}_+, q_i)$ weight spaces

Saitoh [24] derived, on the whole line $\mathbb{R}$, a weight $L_p(\mathbb{R},|q_i|)$ norm inequality related to the Fourier convolution operator, which has the form

$$\left|(P_1q_1)\underset{F}{*}(P_2q_2)(q_1\underset{F}{*}q_2)^{\frac{1}{p}-1}\right| \le \|P_1\|_{L_p(\mathbb{R},|q_1|)}\|P_2\|_{L_p(\mathbb{R},|q_2|)},$$

with $p>1$, where q_i are nonvanishing functions and $P_i\in L_p(\mathbb{R},|q_i|)$, for all $i=1,2$; see also [7] for new convolutions and the corresponding norm inequalities. Here, the norm of P_i in the weight space $L_p(\mathbb{R},|q_i|)$ is understood as

$$\|P_i\|_{L_p(\mathbb{R},|q_i|)} = \left[\int_{-\infty}^{+\infty} |P_i(x_1)|^p |q_i(x_1)|\, dx_1\right]^{\frac{1}{p}}, \quad i = 1, 2.$$

In this section, we develop the analogue of this inequality on the half-line $\mathbb{R}_+$, for the polyconvolution operator (5); accordingly, all the spaces below are weight spaces $L_p(\mathbb{R}_+, q_i)$ over $\mathbb{R}_+$.

Theorem 5. Suppose that q_1, q_2, q_3 are nonvanishing positive functions such that the polyconvolution operator $\overset{\beta}{\underset{F_c,L}{*}}(q_1, q_2, q_3)$ is well-defined. For any $P_i \in L_p(\mathbb{R}_+, q_i)$, with $p > 1$, the following $L_p(\mathbb{R}_+, q_i)$ weight inequality holds:

$$\begin{aligned} &\left\| \overset{\beta}{\underset{F_c,L}{*}} (P_1q_1, P_2q_2, P_3q_3) \overset{\beta}{\underset{F_c,L}{*}} (q_1, q_2, q_3)^{\frac{1}{p}-1} \right\|^p_{L_p(\mathbb{R}_+)} \\ &\quad \le \|P_1\|^p_{L_p(\mathbb{R}_+,q_1)} \|P_2\|^p_{L_p(\mathbb{R}_+,q_2)} \|P_3\|^p_{L_p(\mathbb{R}_+,q_3)}. \end{aligned} \tag{21}$$

Proof. From Definition 1, we have

$$\begin{aligned} &\left\| \overset{\beta}{\underset{F_c,L}{*}} (P_1q_1, P_2q_2, P_3q_3) \overset{\beta}{\underset{F_c,L}{*}} (q_1, q_2, q_3)^{\frac{1}{p}-1} \right\|^p_{L_p(\mathbb{R}_+)} \\ &= \int_0^\infty \left| \overset{\beta}{\underset{F_c,L}{*}} (P_1q_1, P_2q_2, P_3q_3)(x_1)\right|^p \left| \overset{\beta}{\underset{F_c,L}{*}} (q_1, q_2, q_3)(x_1)\right|^{1-p} dx_1 \\ &= \frac{1}{2\pi} \int_0^\infty \Bigg\{ \Bigg| \int_0^\infty \int_0^\infty \int_0^\infty \Bigg[\\ &\qquad \frac{v_1 + w_1}{(v_1 + w_1)^2 + (u_1 + x_1 + 1)^2} + \frac{v_1 + w_1}{(v_1 + w_1)^2 + (u_1 - x_1 + 1)^2} \\ &\qquad + \frac{v_1 + w_1}{(v_1 + w_1)^2 + (u_1 + x_1 - 1)^2} + \frac{v_1 + w_1}{(v_1 + w_1)^2 + (u_1 - x_1 - 1)^2} \Bigg] \\ &\qquad \times (P_1q_1)(u_1)(P_2q_2)(v_1) \times (P_3q_3)(w_1)\, du_1\, dv_1\, dw_1 \Bigg|^p \\ &\quad \times \Bigg| \int_0^\infty \int_0^\infty \int_0^\infty \Bigg[\\ &\qquad \frac{v_1 + w_1}{(v_1 + w_1)^2 + (u_1 + x_1 + 1)^2} + \frac{v_1 + w_1}{(v_1 + w_1)^2 + (u_1 - x_1 + 1)^2} \\ &\qquad + \frac{v_1 + w_1}{(v_1 + w_1)^2 + (u_1 + x_1 - 1)^2} + \frac{v_1 + w_1}{(v_1 + w_1)^2 + (u_1 - x_1 - 1)^2} \Bigg] \\ &\qquad \times q_1(u_1)q_2(v_1)q_3(w_1)\, du_1\, dv_1\, dw_1 \Bigg|^{1-p} \Bigg\} dx_1. \end{aligned} \tag{22}$$

Alternatively, using Hölder's inequality with q the conjugate exponent of p, we get

$$\left| \int_0^\infty \int_0^\infty \int_0^\infty \left[\frac{v_1 + w_1}{(v_1 + w_1)^2 + (u_1 + x_1 + 1)^2} + \frac{v_1 + w_1}{(v_1 + w_1)^2 + (u_1 - x_1 + 1)^2} \right.\right.$$

$$
\begin{aligned}
&+\frac{v_1+w_1}{(v_1+w_1)^2+(u_1+x_1-1)^2}+\frac{v_1+w_1}{(v_1+w_1)^2+(u_1-x_1-1)^2}\Big]\\
&\times (P_1q_1)(u_1)(P_2q_2)(v_1)(P_3q_3)(w_1)\,du_1\,dv_1\,dw_1\Big|\\
\le&\Big\{\int_0^\infty\int_0^\infty\int_0^\infty|\\
&\frac{v_1+w_1}{(v_1+w_1)^2+(u_1+x_1+1)^2}+\frac{v_1+w_1}{(v_1+w_1)^2+(u_1-x_1+1)^2}\\
&+\frac{v_1+w_1}{(v_1+w_1)^2+(u_1+x_1-1)^2}+\frac{v_1+w_1}{(v_1+w_1)^2+(u_1-x_1-1)^2}|\\
&\times|P_1(u_1)|^pq_1(u_1)\\
&\quad\times|P_2(v_1)|^pq_2(v_1)|P_3(w_1)|^pq_3(w_1)\,du_1\,dv_1\,dw_1\Big\}^{\frac{1}{p}}\\
&\times\Big\{\int_0^\infty\int_0^\infty\int_0^\infty|\\
&\frac{v_1+w_1}{(v_1+w_1)^2+(u_1+x_1+1)^2}+\frac{v_1+w_1}{(v_1+w_1)^2+(u_1-x_1+1)^2}\\
&+\frac{v_1+w_1}{(v_1+w_1)^2+(u_1+x_1-1)^2}+\frac{v_1+w_1}{(v_1+w_1)^2+(u_1-x_1-1)^2}|\\
&\times q_1(u_1)q_2(v_1)q_3(w_1)\,du_1\,dv_1\,dw_1\Big\}^{\frac{1}{q}}. \qquad (23)
\end{aligned}
$$

From (22) and (23), we obtain

$$
\begin{aligned}
&\Big\|\overset{\beta}{*}_{F_c,L}(P_1q_1,P_2q_2,P_3q_3)\ \overset{\beta}{*}_{F_c,L}(q_1,q_2,q_3)^{\frac{1}{p}-1}\Big\|^p_{L_p(\mathbb{R}_+)}\\
&\le\frac{1}{2\pi}\int_0^\infty\Big\{\Big[\int_0^\infty\int_0^\infty\int_0^\infty|\\
&\quad\frac{v_1+w_1}{(v_1+w_1)^2+(u_1+x_1+1)^2}+\frac{v_1+w_1}{(v_1+w_1)^2+(u_1-x_1+1)^2}\\
&\quad+\frac{v_1+w_1}{(v_1+w_1)^2+(u_1+x_1-1)^2}+\frac{v_1+w_1}{(v_1+w_1)^2+(u_1-x_1-1)^2}|\\
&\quad\times|P_1(u_1)|^pq_1(u_1)\\
&\qquad\times|P_2(v_1)|^pq_2(v_1)|P_3(w_1)|^pq_3(w_1)\,du_1\,dv_1\,dw_1\Big]\\
&\quad\times\Big[\int_0^\infty\int_0^\infty\int_0^\infty|\\
&\quad\frac{v_1+w_1}{(v_1+w_1)^2+(u_1+x_1+1)^2}+\frac{v_1+w_1}{(v_1+w_1)^2+(u_1-x_1+1)^2}\\
&\quad+\frac{v_1+w_1}{(v_1+w_1)^2+(u_1+x_1-1)^2}+\frac{v_1+w_1}{(v_1+w_1)^2+(u_1-x_1-1)^2}|
\end{aligned}
$$

$$\times q_1(u_1)q_2(v_1)q_3(w_1)\,du_1\,dv_1\,dw_1\Big]^{\frac{p}{q}}\Big\}$$
$$\times\Big[\int_0^\infty\int_0^\infty\int_0^\infty\Big|$$
$$\frac{v_1+w_1}{(v_1+w_1)^2+(u_1+x_1+1)^2}+\frac{v_1+w_1}{(v_1+w_1)^2+(u_1-x_1+1)^2}$$
$$+\frac{v_1+w_1}{(v_1+w_1)^2+(u_1+x_1-1)^2}+\frac{v_1+w_1}{(v_1+w_1)^2+(u_1-x_1-1)^2}\Big|$$
$$\times q_1(u_1)q_2(v_1)q_3(w_1)\,du_1\,dv_1\,dw_1\Big]^{1-p}dx_1.$$

Since $\frac{1}{p}+\frac{1}{q}=1$, we have $\frac{p}{q}+1-p=0$. Therefore, we have

$$\Big\|\overset{\beta}{*}_{F_c,L}(P_1q_1,P_2q_2,P_3q_3)\ \overset{\beta}{*}_{F_c,L}(q_1,q_2,q_3)^{\frac{1}{p}-1}\Big\|^p_{L_p(\mathbb{R}_+)}$$
$$\le\frac{1}{2\pi}\int_0^\infty\int_0^\infty\int_0^\infty\int_0^\infty\Big|$$
$$\frac{v_1+w_1}{(v_1+w_1)^2+(u_1+x_1+1)^2}+\frac{v_1+w_1}{(v_1+w_1)^2+(u_1-x_1+1)^2}$$
$$+\frac{v_1+w_1}{(v_1+w_1)^2+(u_1+x_1-1)^2}+\frac{v_1+w_1}{(v_1+w_1)^2+(u_1-x_1-1)^2}\Big|$$
$$\times|P_1(u_1)|^pq_1(u_1)|P_2(v_1)|^pq_2(v_1)|P_3(w_1)|^pq_3(w_1)\,du_1\,dv_1\,dw_1\,dx_1.$$

From the supposition $P_i\in L_p(\mathbb{R}_+,q_i)$, $i=1,2,3$, using Fubini's theorem, we get

$$\Big\|\overset{\beta}{*}_{F_c,L}(P_1q_1,P_2q_2,P_3q_3)\ \overset{\beta}{*}_{F_c,L}(q_1,q_2,q_3)^{\frac{1}{p}-1}\Big\|^p_{L_p(\mathbb{R}_+)}$$
$$\le\frac{1}{2\pi}\Big[\int_0^\infty\int_0^\infty\int_0^\infty\int_0^\infty\Big|$$
$$\frac{v_1+w_1}{(v_1+w_1)^2+(u_1+x_1+1)^2}+\frac{v_1+w_1}{(v_1+w_1)^2+(u_1-x_1+1)^2}$$
$$+\frac{v_1+w_1}{(v_1+w_1)^2+(u_1+x_1-1)^2}+\frac{v_1+w_1}{(v_1+w_1)^2+(u_1-x_1-1)^2}\Big|dx_1\Big]$$
$$\times\Big[\int_0^\infty|P_1(u_1)|^pq_1(u_1)\,du_1\Big]$$
$$\times\Big[\int_0^\infty|P_2(v_1)|^pq_2(v_1)\,dv_1\Big]\Big[\int_0^\infty|P_3(w_1)|^pq_3(w_1)\,dw_1\Big]$$
$$=\|P_1\|^p_{L_p(\mathbb{R}_+,q_1)}\|P_2\|^p_{L_p(\mathbb{R}_+,q_2)}\|P_3\|^p_{L_p(\mathbb{R}_+,q_3)}.$$

□

5 Applications to solving integral equations, systems of integral equations and an integro-differential equation of Barbashin-type

In this section, we recall the Wiener–Lévy theorem: If f is the Fourier integral transform of a function in $L(\mathbb{R})$, $\Phi(u)$ is an analytic function on the value region of f and $\Phi(0) = 0$. Then $\Phi[f(x)]$ is the Fourier integral transform of a function in $L(\mathbb{R})$.

Remark 1. The Wiener–Lévy theorem also holds when f is the Fourier cosine integral transform of a function in $L(\mathbb{R}_+)$.

We unitize the Banach algebra in which expressions of the form

$$\frac{F_c g}{1 + F_c g}$$

are represented as Fourier-cosine transforms of L^1-functions.

5.1 On the Toeplitz plus Hankel-type integral equation

In this section, we consider the Toeplitz plus Hankel-type integral equation [34]

$$f_1(x_1) + \int_0^M [m_1(x_1 + y_1) + m_2(x_1 - y_1)] f_1(y_1)\, dy_1 = g_1(x_1), \qquad 0 \le x_1, y_1 \le M, \tag{24}$$

where g_1 is a given function, f_1 is an unknown function, and $m_1(x_1 + y_1) + m_2(x_1 - y_1)$ is the Toeplitz plus Hankel-type kernel. Equation (24) has practical applications in many fields such as scattering theory, fluid dynamics, linear filtering theory, problems in radiative wave transmission, and so on. However, up to now, a closed-form solution of (24) for the general case remains an open problem. Therefore, with the help of (5), we study a class of Toeplitz plus Hankel-type integral equations related to this polyconvolution operator.

We consider the following integral equation:

$$f_1(x_1) + \frac{1}{\sqrt{2\pi}} \int_0^\infty f_1(v_1)[g_1(x_1 + v_1) + g_1(|x_1 - v_1|)]\, dv_1 = \frac{1}{2\pi} \int_0^\infty \int_0^\infty \int_0^\infty \Big[$$

$$\frac{v_1+w_1}{(v_1+w_1)^2+(u_1+x_1+1)^2}+\frac{v_1+w_1}{(v_1+w_1)^2+(u_1-x_1+1)^2}$$
$$+\frac{v_1+w_1}{(v_1+w_1)^2+(u_1+x_1-1)^2}+\frac{v_1+w_1}{(v_1+w_1)^2+(u_1-x_1-1)^2}\Big]$$
$$\times h_1(u_1)k_1(v_1)l_1(w_1)\,du_1\,dv_1\,dw_1, \quad x_1>0, \tag{25}$$

where g_1, h_1, k_1, l_1 are functions in $L_1(\mathbb{R}_+)$ and $f_1(x_1)$ is an unknown function in $L_1(\mathbb{R}_+)$.

Theorem 6. Assume that $1+(F_c g_1)(y_1)\neq 0$, for all $y_1>0$. Then the integral equation (25) has a unique solution in $L_1(\mathbb{R}_+)$ in the form

$$f_1(x_1)=\overset{\beta}{\underset{F_c,L}{*}}(h_1,k_1,l_1)(x_1)-\overset{\beta}{\underset{F_c,L}{*}}\big\{(h_1\underset{F_c}{*}p_1),k_1,l_1\big\}(x_1),$$

where $p_1\in L_1(\mathbb{R}_+)$ is defined by

$$(F_c p_1)(y_1)=\frac{(F_c g_1)(y_1)}{1+(F_c g_1)(y_1)}.$$

Proof. Using (3) and (5), equation (25) can be rewritten in the form

$$f_1(x_1)+(f_1\underset{F_c}{*}g_1)=\big[\overset{\beta}{\underset{F_c,L}{*}}(h_1,k_1,l_1)\big](x_1).$$

Because f_1, g_1, $\overset{\beta}{\underset{F_c,L}{*}}(h_1,k_1,l_1)(x_1)$ and $(f_1\underset{F_c}{*}g_1)$ are functions in $L_1(\mathbb{R}_+)$, one can apply the factorization properties (4) and (6) to get

$$(F_c f_1)(y_1)+(F_c f_1)(y_1)(F_c g_1)(y_1)=\cos y_1\,(F_c h_1)(y_1)(Lk_1)(y_1)(Ll_1)(y_1).$$

So

$$(F_c f_1)(y_1)[1+(F_c g_1)(y_1)]=\cos y_1\,(F_c h_1)(y_1)(Lk_1)(y_1)(Ll_1)(y_1).$$

From the condition $1+(F_c g_1)(y_1)\neq 0$, for all $y_1>0$, we have

$$(F_c f_1)(y_1)=\cos y_1\,(F_c h_1)(y_1)(Lk_1)(y_1)(Ll_1)(y_1)\left[1-\frac{(F_c g_1)(y_1)}{1+(F_c g_1)(y_1)}\right].$$

Due to the Wiener–Lévy theorem [1], there exists a function $p_1\in L_1(\mathbb{R}_+)$ such that

$$(F_c p_1)(y_1)=\frac{(F_c g_1)(y_1)}{1+(F_c g_1)(y_1)}.$$

Therefore, from (4) and (6), we get

$$(F_c f_1)(y_1) = F_c\big[\overset{\beta}{\underset{F_c,L}{*}}(h_1,k_1,l_1)\big](y_1) - F_c\big[\overset{\beta}{\underset{F_c,L}{*}}\big\{(h_1 \underset{F_c}{*} p_1),k_1,l_1\big\}\big](y_1).$$

Thus

$$f_1(x_1) = \overset{\beta}{\underset{F_c,L}{*}}(h_1,k_1,l_1)(x_1) - \overset{\beta}{\underset{F_c,L}{*}}\big\{(h_1 \underset{F_c}{*} p_1),k_1,l_1\big\}(x_1) \in L_1(\mathbb{R}_+).$$

□

5.2 The integral equation of polyconvolution type

We consider the following integral equation:

$$\begin{aligned} f_1(x_1) + \frac{1}{2\pi}\int_0^\infty\int_0^\infty\int_0^\infty \Big[& \\ \frac{v_1+w_1}{(v_1+w_1)^2+(u_1+x_1+1)^2} + \frac{v_1+w_1}{(v_1+w_1)^2+(u_1-x_1+1)^2} & \\ + \frac{v_1+w_1}{(v_1+w_1)^2+(u_1+x_1-1)^2} + \frac{v_1+w_1}{(v_1+w_1)^2+(u_1-x_1-1)^2}\Big] & \\ \times f_1(u_1)g_1(v_1)h_1(w_1)\,du_1\,dv_1\,dw_1 = p_1(x_1), \quad x_1>0, & \end{aligned} \tag{26}$$

where g_1, h_1, p_1 are functions in $L_1(\mathbb{R}_+)$ and $f_1(x_1)$ is an unknown function in $L_1(\mathbb{R}_+)$.

Theorem 7. Assume that $1+\cos y_1\,(Lg_1)(y_1)(Lh_1)(y_1) \neq 0$, for all $y_1 > 0$. Then the integral equation (26) has a unique solution in $L_1(\mathbb{R}_+)$ in the form

$$f_1(x_1) = p_1(x_1) - \Big(p_1 \underset{F_c}{*} q_1\Big)(x_1),$$

where $q_1(x_1) \in L_1(\mathbb{R}_+)$ is defined by

$$(F_c q_1)(y_1) = \frac{F_c\big(g_1 \overset{\beta}{\underset{F_c,L}{*}} h_1\big)(y_1)}{1+F_c\big(g_1 \overset{\beta}{\underset{F_c,L}{*}} h_1\big)(y_1)}.$$

Proof. Using (5), equation (26) can be rewritten in the following form:

$$f_1(x_1) + \big[\overset{\beta}{\underset{F_c,L}{*}}(f_1,g_1,h_1)\big](x_1) = p_1(x_1).$$

Because f_1, $\overset{\beta}{\underset{F_c,L}{*}}(f_1,g_1,h_1)$ and p_1 are functions in $L_1(\mathbb{R}_+)$, applying the Fourier cosine integral transform, we obtain

$$(F_c f_1)(y_1) + \cos y_1\,(F_c f_1)(y_1)(Lg_1)(y_1)(Lh_1)(y_1) = (F_c p_1)(y_1).$$

Under the condition $1 + \cos y_1\,(Lg_1)(y_1)(Lh_1)(y_1) \neq 0$, for all $y_1 > 0$, we have

$$(F_c f_1)(y_1) = (F_c p_1)(y_1)\left[1 - \frac{\cos y_1\,(Lg_1)(y_1)(Lh_1)(y_1)}{1 + \cos y_1\,(Lg_1)(y_1)(Lh_1)(y_1)}\right].$$

The generalized convolution operator with the weight function $\beta(y_1) = \cos y_1$ for two functions g_1, h_1 is denoted by $\big(g_1 \overset{\beta}{\underset{F_c,L}{*}} h_1\big)(x_1)$ and defined by

$$\big(g_1 \overset{\beta}{\underset{F_c,L}{*}} h_1\big)(x_1) = \frac{1}{\sqrt{2\pi}} \int_0^\infty \int_0^\infty \left[\frac{u_1 + v_1}{(u_1 + v_1)^2 + (x_1 + 1)^2} + \frac{u_1 + v_1}{(u_1 + v_1)^2 + (x_1 - 1)^2}\right] g_1(u_1) h_1(v_1)\, du_1\, dv_1.$$

We prove that if g_1, h_1 are arbitrary functions in $L_1(\mathbb{R}_+)$, then $\big(g_1 \overset{\beta}{\underset{F_c,L}{*}} h_1\big)(x_1)$ belongs to $L_1(\mathbb{R}_+)$ and satisfies the factorization property

$$F_c\big(g_1 \overset{\beta}{\underset{F_c,L}{*}} h_1\big)(y_1) = \cos y_1\,(Lg_1)(y_1)(Lh_1)(y_1), \qquad \text{for all } y_1 > 0.$$

Therefore, we have

$$(F_c f_1)(y_1) = (F_c p_1)(y_1)\left[1 - \frac{F_c\big(g_1 \overset{\beta}{\underset{F_c,L}{*}} h_1\big)(y_1)}{1 + F_c\big(g_1 \overset{\beta}{\underset{F_c,L}{*}} h_1\big)(y_1)}\right].$$

Due to the Wiener–Lévy theorem [1], there exists a function $q_1 \in L_1(\mathbb{R}_+)$ such that

$$(F_c q_1)(y_1) = \frac{F_c\big(g_1 \overset{\beta}{\underset{F_c,L}{*}} h_1\big)(y_1)}{1 + F_c\big(g_1 \overset{\beta}{\underset{F_c,L}{*}} h_1\big)(y_1)}.$$

Therefore, from (4), we get

$$(F_c f_1)(y_1) = (F_c p_1)(y_1) - F_c\Big(p_1 \underset{F_c}{*} q_1\Big)(y_1).$$

Therefore

$$f_1(x_1) = p_1(x_1) - \Big(p_1 \underset{F_c}{*} q_1\Big)(x_1) \in L_1(\mathbb{R}_+).$$

□

5.3 The system of integral equations of polyconvolution-type

Consider the system of integral equations

$$\begin{aligned} &f_1(x_1) - \frac{1}{2\pi}\int_0^\infty\int_0^\infty\int_0^\infty g_1(u_1)l_1(v_1)h_1(w_1) \\ &\quad\times\Big[\frac{v_1+w_1}{(v_1+w_1)^2+(u_1+x_1+1)^2} + \frac{v_1+w_1}{(v_1+w_1)^2+(u_1-x_1+1)^2} \\ &\quad + \frac{v_1+w_1}{(v_1+w_1)^2+(u_1+x_1-1)^2} + \frac{v_1+w_1}{(v_1+w_1)^2+(u_1-x_1-1)^2}\Big] \\ &\qquad du_1\,dv_1\,dw_1 = p_1(x_1), \\ &\frac{1}{\sqrt{2\pi}}\int_0^\infty f_1(u_1)[r_1(u_1+x_1)+r_1(|u_1-x_1|)]\,du_1 \\ &\quad + g_1(x_1) = q_1(x_1), \quad x_1>0, \end{aligned} \tag{27}$$

where l_1, h_1, r_1, p_1, q_1 are functions in $L_1(\mathbb{R}_+)$ and f_1, g_1 are unknown functions.

Theorem 8. If the following condition is satisfied:

$$1 + F_c\big[\overset{\beta}{\underset{F_c,L}{*}}(r_1,l_1,h_1)\big](y_1) \neq 0, \quad \text{for all } y_1>0,$$

then system (27) has a unique solution in $L_1(\mathbb{R}_+)$ given by

$$\begin{aligned} f_1(x_1) &= p_1(x_1) - \big[\overset{\beta}{\underset{F_c,L}{*}}(q_1,l_1,h_1)\big](x_1) - \big(p_1 \underset{F_c}{*} k_1\big)(x_1) \\ &\quad + \Big\{\big[\overset{\beta}{\underset{F_c,L}{*}}(q_1,l_1,h_1)\big] \underset{F_c}{*} k_1\Big\}(x_1); \\ g_1(x_1) &= q_1(x_1) - \big(r_1 \underset{F_c}{*} p_1\big)(x_1) - \big(q_1 \underset{F_c}{*} k_1\big)(x_1) \\ &\quad + \big[(r_1 \underset{F_c}{*} p_1) \underset{F_c}{*} k_1\big](x_1), \end{aligned}$$

where $k_1(x_1) \in L_1(\mathbb{R}_+)$ is defined by

$$(F_c k_1)(y_1) = \frac{F_c\big[\overset{\beta}{\underset{F_c,L}{*}}(r_1,l_1,h_1)\big](y_1)}{1 + F_c\big[\overset{\beta}{\underset{F_c,L}{*}}(r_1,l_1,h_1)\big](y_1)}.$$

Proof. Using Definition 1 and (3), system (27) can be rewritten in the form

$$\begin{cases} f_1(x_1) - \big[\overset{\beta}{*}_{F_c,L}(g_1,l_1,h_1)\big](x_1) = p_1(x_1), \\ \big(f_1 \underset{F_c}{*} r_1\big)(x_1) + g_1(x_1) = q_1(x_1). \end{cases}$$

Applying the Fourier cosine integral transform to the system, we have

$$\begin{aligned} &(F_c f_1)(y_1) - \cos y_1\,(F_c g_1)(y_1)(Ll_1)(y_1)(Lh_1)(y_1) = (F_c p_1)(y_1), \\ &(F_c f_1)(y_1)(F_c r_1)(y_1) + (F_c g_1)(y_1) = (F_c q_1)(y_1). \end{aligned}$$

The reciprocal of the determinant of this linear algebraic system has the form

$$\frac{1}{\Delta} = 1 - \frac{F_c\big[\overset{\beta}{*}_{F_c,L}(r_1,l_1,h_1)\big](y_1)}{1 + F_c\big[\overset{\beta}{*}_{F_c,L}(r_1,l_1,h_1)\big](y_1)}.$$

Furthermore, according to the Wiener–Lévy theorem [1], there exists a function $k_1 \in L_1(\mathbb{R}_+)$ such that

$$(F_c k_1)(y_1) = \frac{F_c\big[\overset{\beta}{*}_{F_c,L}(r_1,l_1,h_1)\big](y_1)}{1 + F_c\big[\overset{\beta}{*}_{F_c,L}(r_1,l_1,h_1)\big](y_1)}.$$

Hence

$$\frac{1}{\Delta} = 1 - (F_c k_1)(y_1).$$

Therefore, using the factorization properties (4) and (6), we get

$$\begin{aligned} (F_c f_1)(y_1) &= (F_c p_1)(y_1) - F_c\big[\overset{\beta}{*}_{F_c,L}(q_1,l_1,h_1)\big](y_1) - F_c\big(p_1 \underset{F_c}{*} k_1\big)(y_1) \\ &\quad + F_c\Big\{\big[\overset{\beta}{*}_{F_c,L}(q_1,l_1,h_1)\big] \underset{F_c}{*} k_1\Big\}(y_1); \\ (F_c g_1)(y_1) &= (F_c q_1)(y_1) - F_c\big(r_1 \underset{F_c}{*} p_1\big)(y_1) - F_c\big(q_1 \underset{F_c}{*} k_1\big)(y_1) \\ &\quad + F_c\big[(r_1 \underset{F_c}{*} p_1) \underset{F_c}{*} k_1\big](y_1). \end{aligned}$$

It follows that

$$\begin{aligned} f_1(x_1) &= p_1(x_1) - \big[\overset{\beta}{*}_{F_c,L}(q_1,l_1,h_1)\big](x_1) - \big(p_1 \underset{F_c}{*} k_1\big)(x_1) \\ &\quad + \Big\{\big[\overset{\beta}{*}_{F_c,L}(q_1,l_1,h_1)\big] \underset{F_c}{*} k_1\Big\}(x_1) \in L_1(\mathbb{R}_+); \\ g_1(x_1) &= q_1(x_1) - \big(r_1 \underset{F_c}{*} p_1\big)(x_1) - \big(q_1 \underset{F_c}{*} k_1\big)(x_1) \\ &\quad + \big[(r_1 \underset{F_c}{*} p_1) \underset{F_c}{*} k_1\big](x_1) \in L_1(\mathbb{R}_+). \end{aligned}$$

□

5.4 On the integro-differential equation of Barbashin-type

The Barbashin integro-differential equation has applications in many fields such as mathematical physics, radiation propagation, and mathematical biology; see, for example, [2, 3]. The general form of Barbashin's equation is

$$\frac{\partial f_1(x_1,y_1)}{\partial x_1} = \theta(x_1,y_1)f_1(x_1,y_1) + \int_a^b K(x_1,y_1,t_1)f_1(x_1,t_1)\,\mathrm{d}t_1 + \eta(x_1,y_1),$$

where $\theta,\eta : I\times[a,b]\to\mathbb{R}$ are given functions, $K : I\times[a,b]^2\to\mathbb{R}$ is the kernel, and I is a closed set in $\mathbb{R}$ (see [2, 3]). In this subsection, we study the integro-differential equation of Barbashin-type

$$f_1(x_1) + \Big(1-\frac{d^2}{dx_1^2}\Big)\big[\overset{\beta}{*}_{F_c,L}(f_1,g_1,h_1)\big](x_1) = q_1(x_1), \tag{28}$$

where g_1,h_1,q_1 are given functions and f_1 is an unknown function.

Theorem 9. Suppose that g_1,h_1,q_1 belong to $L_2(\mathbb{R}_+)$ and are given functions satisfying the condition: $1+(1+y_1^2)\cos y_1\,(Lg_1)(y_1)(Lh_1)(y_1)\neq 0$;

$$\frac{(F_cq_1)(y_1)}{1+(1+y_1^2)\cos y_1\,(Lg_1)(y_1)(Lh_1)(y_1)} \in L_2(\mathbb{R}_+), \quad y_1>0, \tag{29}$$

and $x_1\cdot\big[\overset{\beta}{*}_{F_c,L}(f_1,g_1,h_1)\big](x_1)$, $x_1^2\cdot\big[\overset{\beta}{*}_{F_c,L}(f_1,g_1,h_1)\big](x_1)\in L_2(\mathbb{R}_+)$, then (28) has a unique solution in $L_2(\mathbb{R}_+)$ which can be presented in the form

$$f_1(x_1) = \sqrt{\frac{2}{\pi}}\int_0^\infty \frac{(F_cq_1)(y_1)\cos(x_1y_1)}{1+(1+y_1^2)\cos y_1\,(Lg_1)(y_1)(Lh_1)(y_1)}\,dy_1, \quad x_1\in\mathbb{R}_+.$$

Proof. Applying the Fourier cosine integral transform to (28) and the factorization identity (6), and using the hypothesis that $x_1\cdot\big[\overset{\beta}{*}_{F_c,L}(f_1,g_1,h_1)\big](x_1)$, $x_1^2\cdot\big[\overset{\beta}{*}_{F_c,L}(f_1,g_1,h_1)\big](x_1)$ belong to $L_2(\mathbb{R}_+)$, we get

$$(F_cf_1)(y_1) + (1+y_1^2)\cos y_1\,(F_cf_1)(y_1)(Lg_1)(y_1)(Lh_1)(y_1) = (F_cq_1)(y_1).$$

From the condition (29), we get

$$(F_c f_1)(y_1) = \frac{1}{1 + (1 + y_1^2)\cos y_1\,(Lg_1)(y_1)(Lh_1)(y_1)}(F_c q_1)(y_1).$$

Using the inverse transform in $L_2(\mathbb{R}_+)$, we get the solution of (28) in $L_2(\mathbb{R}_+)$ and express it as follows:

$$f_1(x_1) = \sqrt{\frac{2}{\pi}}\int_0^\infty \frac{(F_c q_1)(y_1)\cos(x_1 y_1)}{1 + (1 + y_1^2)\cos y_1\,(Lg_1)(y_1)(Lh_1)(y_1)}\,dy_1, \quad x_1 \in \mathbb{R}_+.$$

□

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Akhiezer, N.I. *Lectures on approximation theory*, Science Publishing House, Moscow, 1965. (In Russian.)

[2] Appell, J.M., Kalitvin, A.S. and Zabrejko, P.P. *Partial integral operators and integro-differential equations*, Pure and Applied Mathematics, Marcel Dekker, New York, 2000.

[3] Barbashin, E.A. *Condition for invariance of stability of solution of integro-differential equation*, Izv. Vyssh. Uchebn. Zaved. Mat. 1 (1957) 25–34. (In Russian.)

[4] Castro, L.P., Guerra, R.C. and Tuan, N.M. *New convolutions weighted by Hermite functions and their applications*, Math. Inequal. Appl. 22 (2019) 719–745.

[5] Castro, L.P., Guerra, R.C. and Tuan, N.M. *Convolutions and integral equations weighted by multi-dimensional Hermite functions*, Asian-Eur. J. Math. 15 (2022) Article ID 2250151, 35 pp.

[6] Castro, L.P., Phuong, N.T.H. and Tuan, T. *Qualitative properties and applications of convolutions associated with Hartley and Fourier transforms*, Bol. Soc. Mat. Mex. (3) 32 (2026) Paper No. 54, 46 pp.

[7] Castro, L.P. and Saitoh, S. *New convolutions and norm inequalities*, Math. Inequal. Appl. 15 (2012) 707–716.

[8] Castro, L.P., Silva, A.S. and Tuan, N.M. *New convolutions with Hermite weight functions*, Bull. Iran. Math. Soc. 47 (2021) Suppl. 1, 365–379.

[9] Castro, L.P., Tuan, N.M. and Tuan, T. *Polyconvolutions and their applications to Fredholm and Barbashin-type integral equations*, Quaest. Math. 49 (2026) 335–369.

[10] Churchill, R.V. *Operational mathematics*, McGraw-Hill, New York, 1972.

[11] Debnath, L. and Bhatta, D. *Integral transforms and their applications*, 2nd ed., Chapman & Hall/CRC Press, New York, 2006.

[12] Gakhov, F.D. and Cherskii, Yu.I. *Equations of convolution type*, Nauka, Moscow, 1978. (In Russian.)

[13] Kakichev, V.A. *On the convolution for integral transforms*, Izv. Vyssh. Uchebn. Zaved. Mat. 2 (1967) 48–57. (In Russian.)

[14] Kakichev, V.A. *Polyconvolution*, TPTU, Taganrog, 1997. (In Russian.)

[15] Kakichev, V.A. and Thao, N.X. *A method for constructing generalized integral convolutions*, Izv. Vyssh. Uchebn. Zaved. Mat. 42 (1998) 31–40. (In Russian.)

[16] Khoa, N.M. *On the convolutions of Fourier-type transforms*, Acta Math. Vietnam. 36 (2011) 283–298.

[17] Khoa, N.M. *On a polyconvolution with a weight function for Fourier cosine and Laplace transforms*, Acta Math. Vietnam. 41 (2016) 549–562.

[18] Khoa, N.M. *On the polyconvolution of Hartley integral transforms H_1, H_1, H_2 and integral equation of polyconvolution type*, Far East J. Math. Sci. 132 (2021) 143–158.

[19] Khoa, N.M. *On the polyconvolution operator with a trigonometric weight function for the Hartley integral transforms and applications*, Integral Transforms Spec. Funct. 34 (2023) 861–877.

[20] Khoa, N.M. and Luong, D.X. *On the polyconvolution of Hartley integral transforms H_1, H_2, H_1 and integral equations*, Aust. J. Math. Anal. Appl. 16 (2019) 1–10.

[21] Khoa, N.M. and Thang, T.V. *On the polyconvolution of Hartley integral transforms H_2 and integral equations*, J. Integral Equations Appl. 32 (2020) 171–180.

[22] Khoa, N.M. and Thang, T.V. *On the polyconvolution with weight function $\gamma(y) = \cos y$ for Hartley integral transforms H_1, H_2, H_1 and integral equations*, Ukrainian Math. J. 75 (2023) 652–662.

[23] Napalkov, V.V. *Convolution equations in multidimensional space*, Nauka, Moscow, 1982.

[24] Saitoh, S. *Weighted L_p-norm inequalities in convolutions*, Survey on Classical Inequalities (T.M. Rassias, ed.), Mathematics and its Applications, Springer, Dordrecht, 2000, pp. 225–234.

[25] Sneddon, I.N. *The use of integral transforms*, McGraw-Hill, New York, 1972.

[26] Sogge, C.D. *Fourier integrals in classical analysis*, Cambridge Tracts in Mathematics 105, Cambridge University Press, Cambridge, 1993.

[27] Thao, N.X. and Khoa, N.M. *On the convolution with a weight function for the cosine-Fourier integral transforms*, Acta Math. Vietnam. 29 (2004) 149–162.

[28] Thao, N.X. and Khoa, N.M. *On the generalized convolution with a weight function for the Fourier sine and cosine transforms*, Integral Transforms Spec. Funct. 17 (2006) 673–685.

[29] Thao, N.X., Khoa, N.M. and Anh, P.T.V. *On the polyconvolution for Hartley, Fourier cosine and Fourier sine transforms*, Integral Transforms Spec. Funct. 24 (2013) 517–531.

[30] Thao, N.X., Tuan, T. and Huy, L.X. *The Fourier-Laplace generalized convolutions and applications to integral equations*, Vietnam J. Math. 41 (2013) 451–464.

[31] Thao, N.X., Tuan, T. and Huy, L.X. *The generalized convolutions with a weight function for Laplace transform*, Nonlinear Funct. Anal. Appl. 19 (2014) 61–77.

[32] Thao, N.X., Tuan, V.K. and Khoa, N.M. *A generalized convolution with a weight function for the Fourier cosine and sine transform*, Fract. Calc. Appl. Anal. 7 (2004) 323–337.

[33] Titchmarsh, E.C. *Introduction to the theory of Fourier integrals*, 3rd ed., Chelsea Publishing Co., New York, 1986.

[34] Tsitsiklis, J.N. and Levy, B.C. *Integral equations and resolvents of Toeplitz plus Hankel kernel*, Report No. LIDS-P 1170, Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, 1981.

[35] Tuan, T. *New polyconvolution product for Fourier-cosine and Laplace integral operators and their applications*, Math. Methods Appl. Sci. 47 (2024) 1680–1700.

[36] Tuan, V.K. and Saigo, M. *Convolution of Hankel transform and its applications to an integral involving Bessel function of first kind*, Int. J. Math. Math. Sci. 18 (1995) 545–550.

[37] Yakubovich, S.B. and Britvina, L.E. *Convolution related to the Fourier and Kontorovich-Lebedev transforms revisited*, Integral Transforms Spec. Funct. 21 (2010) 255–276.

Iranian Journal of Numerical Analysis and Optimization
Vol. 17, No. 1, 2027, pp 273–289
https://doi.org/10.22067/ijnao.2026.97045.1806
https://ijnao.um.ac.ir/

CC BY NC

Research Article

Mathematical modeling and optimal control of hepatitis B virus dynamics

L. El Houari*, M. Elkaf and J. Danane

Abstract

In this study, we examine the dynamics of Hepatitis B Virus, formulated through a system of four differential equations that describe the interactions between healthy hepatocytes, infected hepatocytes, capsids, and HBV. To establish the well-posedness of the proposed model, we prove the positivity and boundedness of solutions under two control strategies. Then, we give some numerical results confirming that control strategies play a crucial role in reducing viral load, leading to improved disease management and prevention of complications such as liver damage. The overall findings emphasize the importance of intervention in maintaining biological balance and enhancing health outcomes.

AMS subject classifications (2020): Primary 92C50, 49J15; Secondary 34A34, 34D20, 65L06, 37N25.

*Corresponding author

Received 17 December 2025; revised 3 July 2026; accepted 8 August 2026

Loubna El Houari
LaMAO Laboratory, Faculty of Sciences, Mohammed First University, Oujda, Morocco. e-mail: l.elhouari@ump.ac.ma.

Mariem Elkaf
Royal Naval School, Casablanca, Morocco. e-mail: elkaf.mariem@gmail.com.

Jaouad Danane
National School of Applied Sciences, Hassan First University, Berrechid, Morocco. e-mail: jaouad-danane@gmail.com

How to cite this article
El Houari, L., Elkaf, M. and Danane, J., Mathematical modeling and optimal control of hepatitis B virus dynamics. *Iran. J. Numer. Anal. Optim.*, 2027; 17(1): 273-289. https://doi.org/10.22067/ijnao.2026.97045.1806

Keywords: HBV, Dynamical system, Ordinary differential equations, Stability, optimal control, Optimization, Numerical simulation.

1 Introduction

The hepatitis B virus (HBV) is a pathogen known to cause serious liver disease by infecting healthy hepatocytes. This condition is highly contagious and can spread through contact with infected bodily fluids [7, 16]. Due to its ease of transmission, HBV infection represents a major global health concern. According to estimates from the World Health Organization (WHO), hepatitis B and C lead to approximately one million deaths annually, with around three million new infections reported each year [19, 20].

To gain a deeper understanding of the dynamics behind hepatitis B viral infection, numerous mathematical models have been developed, introducing the capsid as an element in the dynamics of HBV disease to show that optimized drug treatments can reduce viral load and increase healthy cells [9, 8, 14]. These models aim to study and predict the virus's behavior and progression. It is well known that infected liver cells release HBV DNA-containing capsids in the form of mature viruses [17, 12, 11]. A significant body of research has focused on optimizing the control of HBV infection by targeting these capsids [6, 13], showing that optimal therapy effectively controls viral replication and improves treatment outcomes. Furthermore, other studies have explored this infection through different mathematical frameworks, including fractional differential equations (FDEs) [18, 1, 2, 5], ordinary differential equations (ODEs) [17, 6, 11], and partial differential equations (PDEs) [4, 10].

More recent investigations [16, 9, 14] have incorporated the capsid parameter into their models, examining aspects such as stability and optimal control strategies. These numerous contributions highlight the crucial role that mathematical modeling plays in enhancing our understanding of the HBV infection process.

Based on a comprehensive review and critical analysis of previous research [11, 15], we propose the following model that consists of four variables describing the dynamics of HBV infection. The problem is formulated as a nonlinear differential system:

$$\begin{cases} \frac{dH}{dt} = s - \mu H - k\frac{HV}{H+I}, \\ \frac{dI}{dt} = k\frac{HV}{H+I} - \delta I, \\ \frac{dD}{dt} = aI - (\beta+\delta)D, \\ \frac{dV}{dt} = \beta D - uV. \end{cases} \tag{1}$$

The variables of our model are as follows:

- (H) Healthy hepatocyte cells,
- (I) Infected cells,
- (D) Capsids,
- (V) Hepatitis B virus.

They vary as follows:

- The hepatocytes (H) are produced at a rate of s and die at a rate of μ. These healthy cells become infected by the virus at a rate of $k\frac{HV}{H+I}$, where k is the infection rate.
- The infected cells (I) arise from the infection of healthy hepatocytes at a rate of $k\frac{HV}{H+I}$ and naturally die at a rate of δ.
- The capsids (D) are produced at a rate of a, transmitted to the blood at a rate of β, and decay at a rate of δ.
- The virus (V) grows in the bloodstream at a rate of β due to DNA-containing capsids and decays at a rate of u.

The parameters of model (1), along with their corresponding units and values, are grouped in Table 1.

To provide clearer insight into this formulation of HBV infection dynamics, we present Figure 1, which illustrates the viral progression and models the various interactions among the variables.

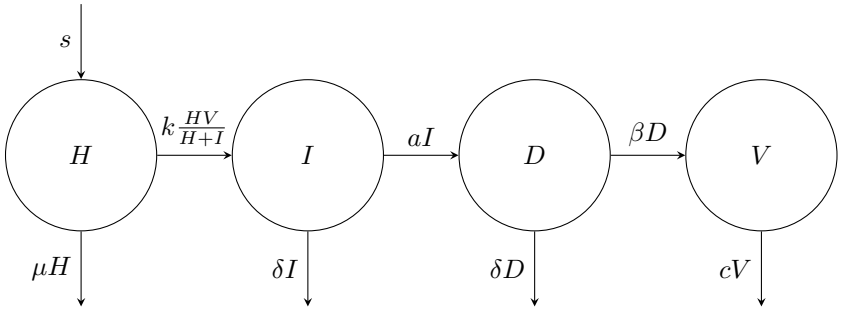


Figure 1: A graphical diagram illustrating the dynamics of HBV.

However, certain therapies or drugs can reduce HBV infection, highlighting the importance of studying the impact of optimal therapeutic control on HBV viral dynamics. For this purpose, we denote α_1 and α_2 as the efficacy rates of drugs or therapies 1 and 2, respectively. Thus, the model becomes

$$\begin{cases} \frac{dH}{dt} = s - \mu H - k(1-\alpha_1)\frac{HV}{H+I}, \\ \frac{dI}{dt} = k(1-\alpha_1)\frac{HV}{H+I} - \delta I, \\ \frac{dD}{dt} = a(1-\alpha_2)I - (\beta+\delta)D, \\ \frac{dV}{dt} = \beta D - uV. \end{cases} \tag{2}$$

Remark 1. The efficacy values of drugs or therapies range between 0 and 1.

The manuscript is organized into several sections. In Section 2, we focus on the stability criterion of the model. Section 3 addresses the model's optimization analysis as previously employed in [3]. In Section 4, we develop a numerical algorithm and present several numerical simulations. Finally, we summarize our findings and discuss future perspectives in the conclusion Section 5.

2 Optimization criterion

In the present section, we analyze the well-posedness of the HBV dynamics model by ensuring the positivity and boundedness of solutions, thereby preserving the biological relevance of the infection model. We also determine the disease's steady states and then examine the threshold for effective therapy.

2.1 Positivity and boundedness

We assume that the initial conditions of the solutions are positive to maintain biological relevance. In this subsection, we show the well-posedness of the solution to model (2).

Theorem 1. The solutions of the given problem in (2) are bounded and nonnegative.

Proof. Indeed, for $(H(t), I(t), D(t), V(t)) \in \mathbb{R}^4_+$, we have

$\frac{dH}{dt}\big|_{H=0} = s \geq 0,$

$\frac{dI}{dt}\big|_{I=0} = k(1-\alpha_1)V \geq 0,$

$\frac{dD}{dt}\big|_{D=0} = a(1-\alpha_2)I \geq 0,$

$\frac{dV}{dt}\big|_{V=0} = \beta D \geq 0.$

Therefore, all solutions initiating in $\mathbb{R}^4_+$ are positive, which shows the nonnegativity of the solution.

To demonstrate that the solutions remain bounded, we obtain the following inequalities from the equations of model (2):

$\frac{dH}{dt} + \mu H \leq s,$
$\frac{d(H+I)}{dt} + \mu_1(H+I) \leq s, \quad$ where we set $\mu_1 = \min(\mu, \delta),$
$\frac{dD}{dt} + (\beta+\delta)D \leq a(1-\alpha_2)I,$
$\frac{dV}{dt} + uV \leq \beta D.$

Thus

$H \leq \left(H_0 - \frac{s}{\mu}\right) e^{-\mu t} + \frac{s}{\mu},$
$H + I \leq \left(H_0 + I_0 - \frac{s}{\mu_1}\right) e^{-\mu_1 t} + \frac{s}{\mu_1}.$

We deduce the boundedness of H. Since H is bounded, I is also bounded.

Thus

$D \leq D_0 e^{-(\beta+\delta)t} + \int_0^t a(1-\alpha_2) I e^{-(\beta+\delta)(t-\zeta)} d\zeta.$

Since I is bounded, D is also bounded.

Thus

$V \leq V_0 e^{-ut} + \int_0^t \beta D e^{-u(t-\zeta)} d\zeta.$

Since D is bounded, V is also bounded. □

2.2 Steady states

Before examining the stability of the viral dynamics, we must calculate the basic reproduction number (BRN). This value represents the average number of new cases generated by a single infected cell. Mathematically, it is defined as the spectral radius of the next-generation matrix.

Using the dominant eigenvalue, we obtain the following calculation results, allowing us to express R_0 as follows:

$$R_0 = \frac{k(1-\alpha_1)(1-\alpha_2)a\beta}{\delta u(\beta+\delta)}.$$

We seek to determine all possible equilibrium points. Simple calculations show that there is one disease-free equilibrium $E_f(\frac{s}{\mu}, 0, 0, 0)$. Furthermore, we can find an endemic steady state $E_e(H_e, I_e, D_e, V_e)$ when

$H_e = \frac{s}{\delta(R_0-1)+\mu}$,
$I_e = \frac{s(R_0-1)}{\delta(R_0-1)+\mu}$,
$D_e = \frac{a(1-\alpha_2)s(R_0-1)}{(\beta+\delta)(\delta(R_0-1)+\mu)}$,
$V_e = \frac{\beta a(1-\alpha_2)s(R_0-1)}{u(\beta+\delta)(\delta(R_0-1)+\mu)}$.

Remark 2. We observe that the endemic equilibrium point exists when R_0 is greater than one.

Theorem 2. The disease-free equilibrium point E_f is locally stable when $R_0 < 1$.

The endemic equilibrium point E_e exists and is locally stable when $R_0 \geq 1$.

Proof. We compute the Jacobian matrix of system (2) at the equilibrium points and discuss the signs of its eigenvalues. **Stability of the disease-free equilibrium E_f:**
The Jacobian matrix at $E_f = \left(\frac{s}{\mu}, 0, 0, 0\right)$ is

$$J(E_f) = \begin{pmatrix} -\mu & 0 & 0 & -k \\ 0 & -\delta & 0 & k \\ 0 & a & -(\beta+\delta) & 0 \\ 0 & 0 & \beta & -u \end{pmatrix}.$$

Its characteristic polynomial is

$$P_{E_f}(\lambda) = (\mu+\lambda)(\delta+\lambda)\left[(\beta+\delta+\lambda)(u+\lambda) - ak\beta\right].$$

The first two eigenvalues are clearly negative. The rest of the roots have negative real parts if and only if

$$(\beta+\delta)u - ak\beta > 0,$$

which is equivalent to

$$R_0 < 1.$$

Hence, E_f is locally asymptotically stable when $R_0 < 1$.
Stability of the endemic equilibrium E_e:
For $R_0 \geq 1$, the equilibrium $E_e(H_e, I_e, D_e, V_e)$ exists.
The Jacobian matrix at $E_e = (H_e, I_e, D_e, V_e)$ is

$$J(E_e) = \begin{pmatrix} -\mu - k\frac{V_e I_e}{(H_e+I_e)^2} - k\frac{V_e}{H_e+I_e} & -k\frac{H_e V_e}{(H_e+I_e)^2} & 0 & -k\frac{H_e}{H_e+I_e} \\ k\frac{V_e I_e}{(H_e+I_e)^2} + k\frac{V_e}{H_e+I_e} & k\frac{H_e V_e}{(H_e+I_e)^2} - \delta & 0 & k\frac{H_e}{H_e+I_e} \\ 0 & a & -(\beta+\delta) & 0 \\ 0 & 0 & \beta & -u \end{pmatrix}.$$

Its characteristic polynomial is

$$P(\lambda) = [(\lambda - a_{11})(\lambda - a_{22}) - a_{12}a_{21}](\lambda + \beta + \delta)(\lambda + u),$$

where
$a_{11} = -\mu - k\frac{V_e I_e}{(H_e+I_e)^2} - k\frac{V_e}{H_e+I_e}$,
$a_{22} = k\frac{H_e V_e}{(H_e+I_e)^2} - \delta$,
$a_{12} = -k\frac{H_e V_e}{(H_e+I_e)^2}$,
and $a_{21} = k\frac{V_e I_e}{(H_e+I_e)^2} + k\frac{V_e}{H_e+I_e}$.

The first two eigenvalues are clearly negative. The rest of the roots have negative real parts if and only if

$$-a_{11} - a_{22} > 0 \quad \text{and} \quad a_{11}a_{22} - a_{12}a_{21} > 0,$$

which is always true under the hypothesis $\delta \geq \mu$. Therefore, all eigenvalues of $J(E_e)$ have negative real parts, and the endemic equilibrium is locally asymptotically stable when $R_0 \geq 1$. □

2.3 Critical therapy efficacy

In system (2), the effectiveness of drugs 1 and 2 is represented by the terms $1-\alpha_1$ and $1-\alpha_2$, respectively. Thus, the values $\alpha_i = 0$ and $\alpha_i = 1$ (for $i = 1, 2$) indicate complete ineffectiveness and full effectiveness of the respective drug. For simplicity, we combine the effectiveness of both drugs into a single term that reflects overall efficacy: $1 - \eta = (1-\alpha_1)(1-\alpha_2)$.

The new BRN becomes

$$R_0 = \frac{k(1-\eta)a\beta}{\delta u(\beta+\delta)}.$$

To determine the critical drug and therapy efficacy η_{crit}, we set $R_0 = 1$.

Thus,

$$\eta_{\text{crit}} = 1 - \frac{\delta u(\beta+\delta)}{ka\beta}.$$

Therefore, for a comprehensive interpretation, if $\eta \leq \eta_{\text{crit}}$ $(R_0 \geq 1)$, then the HBV infection will continue to persist.

Conversely, if $\eta > \eta_{\text{crit}}$ $(R_0 < 1)$, then this suggests disease elimination.

3 Optimal control techniques

We formulate the optimal control problem by introducing the cost functional J, defined by the following expression:

$$J(\alpha_1, \alpha_2) = \frac{1}{2} \int_0^{t_f} \left\{ Q_1 \alpha_1^2(t) + Q_2 \alpha_2^2(t) + Q_3 V^2(t) + Q_4 D^2(t) \right\} dt, \tag{3}$$

where t_f denotes the duration of the treatment. The positive constants Q_1, Q_2, Q_3, and Q_4 represent the benefit-cost weights associated with the control functions α_1 and α_2, respectively. The control functions $\alpha_1(t)$ and $\alpha_2(t)$ are assumed to be bounded and Lebesgue integrable over the time interval $[0, t_f]$. The aim is to minimize the objective functional defined in (3) by simultaneously reducing the number of infected cells, lowering the viral load, and minimizing the cost associated with treatment. In this context, we construct the Hamiltonian associated with the optimal control problem in order to derive the necessary optimality conditions.

More precisely, for the state vector $x(t) = (H, I, D, V)$, where H, I, D, and V denote the healthy cells, infected cells, capsids, and virus, respectively, the Hamiltonian is given by

$$\mathcal{H} = \frac{1}{2}(Q_1 \alpha_1^2 + Q_2 \alpha_2^2 + Q_3 V^2 + Q_4 D^2) + \lambda_1 \frac{\partial H}{\partial t} + \lambda_2 \frac{\partial I}{\partial t} + \lambda_3 \frac{\partial D}{\partial t} + \lambda_4 \frac{\partial V}{\partial t}. \tag{4}$$

The Hamiltonian combines the objective functional with the state equations through the adjoint variables. The we seek an optimal control pair

$$(\alpha_1^*, \alpha_2^*) = \underset{U}{argmin}\, \mathcal{H}(x^*, \alpha_1, \alpha_2, \lambda_1, \lambda_2, \lambda_3, \lambda_4, t) \tag{5}$$

that achieves the desired therapeutic outcomes in an efficient and cost-effective manner.
The functional $\mathcal{H}$ is Fréchet differentiable with respect to its arguments. Moreover, we assume that $(\alpha_1, \alpha_2) \in U$, where U denotes the control set defined by

$$\begin{aligned} U = \{&(\alpha_1(t), \alpha_2(t)) \text{ such that } \alpha_1(t), \alpha_2(t) \text{ are measurable} \\ &\text{and satisfy } 0 \leq \alpha_i(t) \leq 1, t \in [t_0, t_f], \text{ for } i = 1, 2\}. \end{aligned} \tag{6}$$

For a minimization problem, Pontryagin's maximum principle gives the condition:

$$\mathcal{H}(x^*, \alpha_1^*, \alpha_2^*, \lambda_1^*, \lambda_2^*, \lambda_3^*, \lambda_4^*, t) = 0. \tag{7}$$

This means that, for fixed $x^*(t), \lambda_1(t), \lambda_2(t), \lambda_3(t), \lambda_4(t)$, and t, the function $x \to \mathcal{H}(x^*, \lambda_1, \lambda_2, \lambda_3, \lambda_4)$ must attain its minimum on the admissible control set U.

3.1 Existence and uniqueness of an optimal control pair

The discussion of the existence, uniqueness, and characterization of the optimal control pair is presented in this section by proving that the Hamiltonian $\mathcal{H}$ admits a unique minimum with respect to the control variable.

The results are established using the following lemmas.

Lemma 1. The functional $\mathcal{H}$ defined in (4) is continuous, strictly convex on the set U, and satisfies

$$\lim_{\|(\alpha_1,\alpha_2)\|\to+\infty} \mathcal{H}(\alpha_1, \alpha_2) = +\infty.$$

Proof. Since it is bounded from below by a quadratic function that goes to infinity, $\mathcal{H}$ tends to ∞, clearly satisfying the properties outlined in Lemma 1 on the set U. Moreover, $\mathcal{H}$ is strictly convex because its second derivative is positive. □

Lemma 2. It holds that U defined in (6) is a nonempty, closed, and convex subset of $L^2([t_0, t_f])$.

Proof. By the definition, the set U of admissible controls and associated state variables is a nonempty, convex, and closed subset. □

In this subsection, we have shown that the functional $\mathcal{H}$ is continuous, strictly convex on the set U, and satisfies

$$\lim_{\|(\alpha_1,\alpha_2)\|\to+\infty} \mathcal{H}(\alpha_1, \alpha_2) = +\infty,$$

where U is a nonempty, convex, and closed set. As a result, we conclude that the functional $\mathcal{H}$ admits a unique minimum (α_1^*, α_2^*).

3.2 Derivation of adjoint equations

The adjoint equations are obtained by differentiating the Hamiltonian with respect to the state variables. More precisely, for each state variable $x_i(t)$, the corresponding adjoint variable $\lambda_i(t)$

satisfies

$$\begin{aligned}
\overset{\bullet}{\lambda}_1 &= \frac{d\lambda_1}{dt} = -\frac{\partial \mathcal{H}}{\partial H} = \lambda_1\mu + (\lambda_1 - \lambda_2)k(1-\alpha_1)\frac{VI}{(H+I)^2},\\
\overset{\bullet}{\lambda}_2 &= \frac{d\lambda_2}{dt} = -\frac{\partial \mathcal{H}}{\partial I} = (\lambda_2 - \lambda_1)(1-\alpha_1)\frac{kHV}{(H+I)^2} + \lambda_2\delta - a(1-\alpha_2)\lambda_3,\\
\overset{\bullet}{\lambda}_3 &= \frac{d\lambda_3}{dt} = -\frac{\partial \mathcal{H}}{\partial D} = \lambda_3(\beta+\delta) - \lambda_4\beta - Q_4 D,\\
\overset{\bullet}{\lambda}_4 &= \frac{d\lambda_4}{dt} = -\frac{\partial \mathcal{H}}{\partial V} = (\lambda_1 - \lambda_2)k(1-\alpha_1)\frac{H}{H+I} + \lambda_4\mu - Q_3 V,
\end{aligned} \tag{8}$$

with $\lambda_1(t_f) = \lambda_2(t_f) = \lambda_3(t_f) = \lambda_4(t_f) = 0$

3.3 Optimality conditions

Assuming that (α_1^*, α_2^*) is an optimal control pair minimizing the Hamiltonian $\mathcal{H}$, we derive below the optimality conditions corresponding to problem (5).

Theorem 3. There exists a unique optimal control pair (α_1^*, α_2^*) that solves the optimization problem (4) such that the gradient of the Hamiltonian $\mathcal{H}$ satisfies the following optimality condition:

$$\begin{aligned}
\alpha_1^* &= (\lambda_2 - \lambda_1)\frac{kHV}{(H+I)Q_1},\\
\alpha_2^* &= \frac{\lambda_3 aI}{Q_2}.
\end{aligned}$$

Proof. Let us assume that (α_1^*, α_2^*) is an optimal control pair minimizing the Hamiltonian $\mathcal{H}$. The optimality is characterized by the first-order necessary condition:

$$\frac{\partial \mathcal{H}}{\partial \alpha_1} = Q_1\alpha_1 + (\lambda_1 - \lambda_2)\frac{kHV}{H+I} = 0,$$

$$\frac{\partial \mathcal{H}}{\partial \alpha_2} = Q_2\alpha_2 - \lambda_3 aI = 0.$$

That is,

$$\alpha_1^* = (\lambda_2 - \lambda_1)\frac{kHV}{(H+I)Q_1},$$

$$\alpha_2^* = \frac{\lambda_3 a I}{Q_2}.$$

□

4 Numerical section

This section is devoted to presenting numerical simulations that illustrate the theoretical results derived in the previous section.

4.1 Iterative algorithm

The optimization problem (5) can be solved using a descent method described by the following algorithm.

1. Select an initial pair (α_1^0, α_2^0);

2. Solve state equations (2) forward in time using implicit Euler method.

3. Solve adjoint equations (8) backward in time using implicit Euler method.

4. Update the controls using the optimality condition, that is, $(\alpha_1^{n+1}, \alpha_2^{n+1}) = (\alpha_1^n, \alpha_2^n) - m_n d_n$, where d_n is the search direction. The search direction d_n is computed depending on the applied optimization technique. Herein, we use $d_n = \nabla \mathcal{H}$,
   Also, m_n (step size) is chosen by using the Armijo rule.

5. The stopping criterion for the minimization process is defined by

$$\|\nabla \mathcal{H}\| < \varepsilon,$$

   where ε is a prescribed convergence parameter.

4.2 Numerical simulations

To solve the system (2) numerically, we use the previous algorithm. Table 1 describes the parameter values used for the simulations [8, 14].

Table 1: Parameter description of system (1)

Parameters	Descriptions	Units	Values
s	Production rate of healthy hepatocytes	cells.mm^{-3}.day^{-1}	2.6×10^7
μ	Natural death rate of healthy hepatocytes	day^{-1}	0.01
k	Infection rate of healthy cells by the virus	cells.virion^{-1}.day^{-1}	3×10^{-4}
δ	Natural death rate of infected cells and capsids	day^{-1}	0.053
a	Production rate of capsids by infected cells	day^{-1}	150
β	Rate at which capsids are converted into virus	day^{-1}	0.87
u	Decay rate of free virus	day^{-1}	3.8

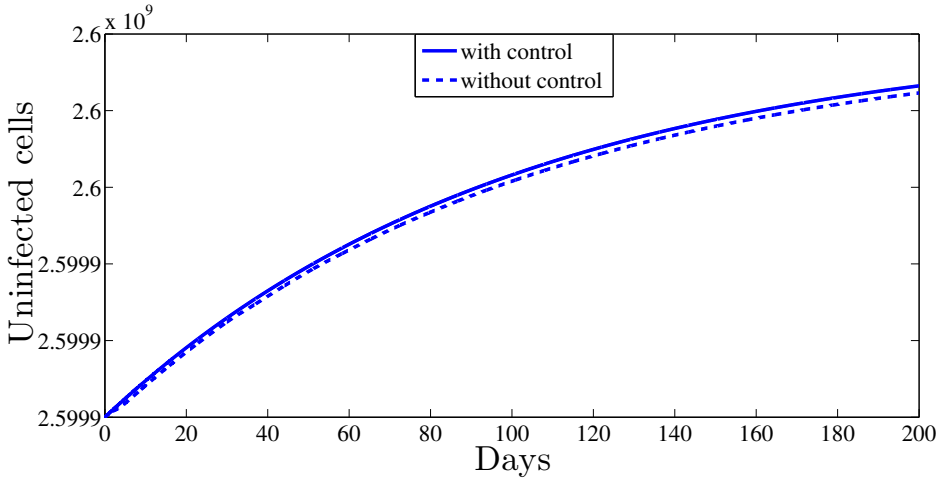


Figure 2: The effect of the control strategy on uninfected hepatocytes

Figure 2 illustrates the dynamics of uninfected hepatocytes over a period of 200 days. A solid blue line in the plot indicates a scenario with therapeutic control, while a dashed blue line indicates a situation without control. The results indicate that the presence of a control strategy leads to a slightly higher number of uninfected hepatocytes over time, suggesting that therapeutic intervention contributes to the preservation of healthy liver cells.

The dynamics of infected hepatocytes throughout time are shown in Figure 3. The number of infected cells is high at first but quickly drops, especially when a control mechanism is in place. Two situations are contrasted in the plot: One with control (solid line) and the other without

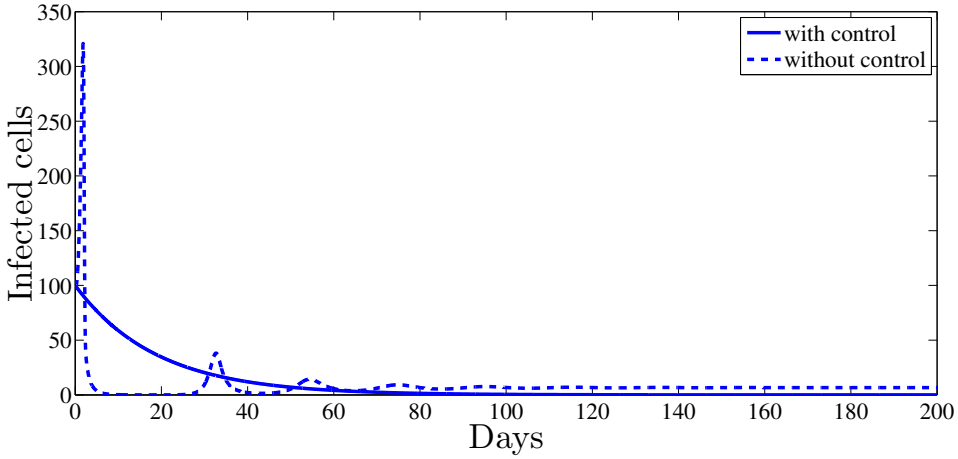


Figure 3: The effect of the control strategy on infected hepatocytes

(dashed line). The outcomes demonstrate the therapeutic efficacy of the control technique by showing that it considerably lowers the number of infected hepatocytes more efficiently than in the uncontrolled scenario.

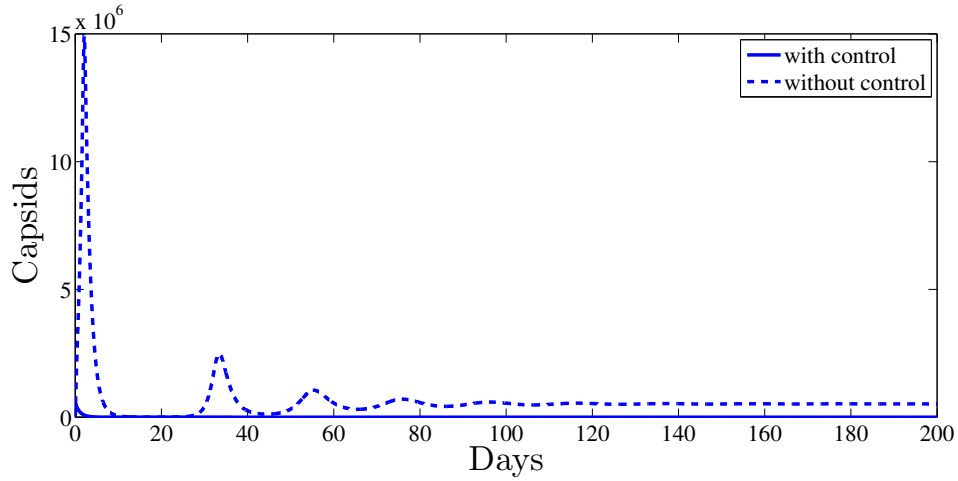


Figure 4: The effect of the control strategy on Capsids

Figure 4 compares the levels of capsids over a 200-day period under two conditions: with control and without control. It likely demonstrates that the control strategy effectively reduces or stabilizes capsid levels over time, in contrast to the uncontrolled condition where capsid levels may rise or persist. The line with control would show a decline or stabilization in capsid levels, while the line without control might indicate higher or increasing levels. This suggests that the

control strategy plays an important role in managing capsid levels, which is critical in fields like virology and drug development.

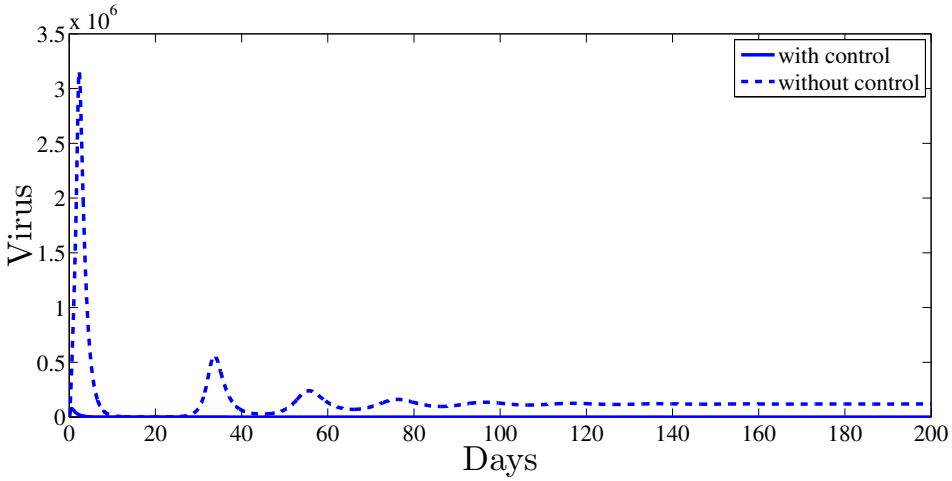


Figure 5: The effect of the control strategy on HBV

Figure 5 compares the levels of HBV (hepatitis B virus) under two conditions. Under the controlled condition, the viral load shows a significant decline or stabilization over time, indicating that the control strategy (such as antiviral therapy or immune modulation) effectively suppresses HBV replication. In contrast, the uncontrolled condition shows that the viral load remains persistently high or increases, signifying unchecked viral replication in the absence of intervention. The curve may fluctuate but remains elevated (e.g., near 2.5×10^6), highlighting the need for control measures. This stark contrast emphasizes the efficacy of the intervention, which is critical for managing infections and preventing liver damage.

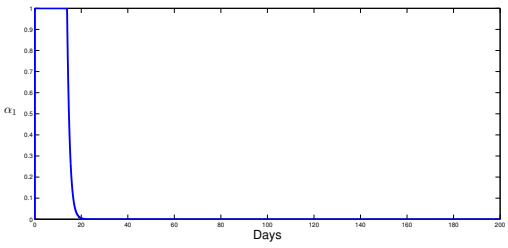


Figure 6: First control

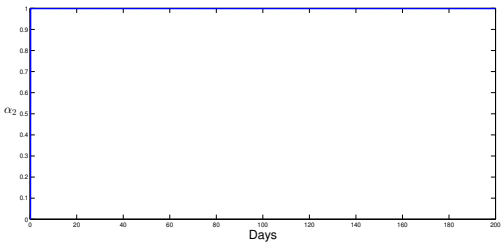


Figure 7: Second control

Figures 2, 3, 4, and 5 visually underscore the importance of control strategies in HBV management, with clear evidence of reduced viral load under treatment. Further details on

the specific intervention (such as drug type or dosage) would enhance the understanding of its impact.

Additionally, we observe in Figures 6 and 7 that the first control should be prioritized over the second within the first 20 days, based on the behavior of both controls. The prioritization of the first control within the first 20 days is consistent with the need to counteract rapid early viral replication. Early intervention via the first control is therefore critical to limit viral spread before secondary controls gain relevance.

5 Conclusion

In this paper, we have studied the HBV dynamics mathematically and numerically, as illustrated by system (2). We started by demonstrating the boundedness and positivity of the solutions, considering two control strategies. We calculated the BRN R_0, the steady states (both disease-free and endemic), and the critical therapy efficacy. Next, we applied optimal control techniques to minimize the proposed objective function. Numerically, we presented our results, confirming that control measures are essential for reducing viral load, which enhances disease management and helps prevent complications such as liver disease. In future work, we propose several extensions to improve the present paper. The current model does not account for several biologically relevant factors, including immune response, drug resistance, and spatial heterogeneity. To address these limitations and open new research avenues, we suggest exploring the development of a stochastic version of the model, a fractional-order formulation, multi-drug optimization approaches, or clinical validation studies.

Acknowledgements

The authors are grateful to their anonymous referees and the editor for their constructive comments.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Ait Ichou, M., Bachraoui, M., Hattaf, K., and Yousfi, N. *Dynamics of a fractional optimal control HBV infection model with capsids and CTL immune response,* Math. Model. Comput. 10(1) (2023) 239–244.

[2] Bachraoui, M., Hattaf, K., and Yousfi, N. *Dynamics of a fractional order HBV infection model with capsids and CTL immune response,* Commun. Math. Biol. Neurosci., 2019 (2019), Article ID 6.

[3] Boudjaj, L., Naji, A., and Ghafrani, F. *Solving biharmonic equation as an optimal control problem using localized radial basis functions collocation method,* Eng. Anal. Bound. Elem. 107 (2019) 208–217.

[4] Churkin, A., Lewkiewicz, S., Reinharz, V., Dahari, H. and Barash, D. *Efficient methods for parameter estimation of ordinary and partial differential equation models of viral hepatitis kinetics,* Mathematics 8(9) (2020) 1483.

[5] Danane, J., Allali, K., and Hammouch, Z. *Mathematical analysis of a fractional differential model of HBV infection with antibody immune response,* Chaos Solitons Fractals 136 (2020) 109787.

[6] Danane, J., Meskaf, A., and Allali, K. *Optimal control of a delayed hepatitis B viral infection model with HBV DNA-containing capsids and CTL immune response,* Optim. Control Appl. Methods 39(3) (2018) 1262–1272.

[7] di Filippo Villa, D., and Navas, M.-C. *Vertical transmission of hepatitis B virus—An update.* Microorganisms 11(5) (2023) 1140.

[8] Elkaf, M., Meskaf, A., and Allali, K. *Mathematical Modeling of HBV Infection with DNA-Containing Capsids and Therapy,* In Nonlinear Dynamics and Complexity: Mathematical Modelling of Real-World Problems (pp. 413-427). Cham: Springer International Publishing, 2022.

[9] Foko, S. *Dynamical analysis of a general delayed HBV infection model with capsids and adaptive immune response in presence of exposed infected hepatocytes,* J. Math. Biol. 88(6) (2024) 75.

[10] Li, M., and Zu, J. *The review of differential equation models of HBV infection dynamics,* J. Virol. Methods 266 (2019) 103–113.

[11] Manna, K. *Global properties of an HBV infection model with HBV DNA-containing capsids and CTL immune response,* Int. J. Appl. Comput. Math. 3 (2017) 2323–2338.

[12] Manna, K., and Chakrabarty, S. P. *Chronic hepatitis B infection and HBV DNA-containing capsids: Modeling and analysis.* Commnu Nonlinear Sci. Numer. Simul. 22(1–3) (2015) 383–395.

[13] Meskaf, A., and Allali, K. *Optimal control of a delayed hepatitis B viral infection model with DNA-containing capsids and cure rate,* Trends in Biomathematics: Mathematical Modeling for Health, Harvesting, and Population Dynamics: Selected works presented at the BIOMAT Consortium Lectures, Morocco 2018 (2019): 23-33.

[14] Meskaf, A., Elkaf, M., and Aamoum, I. *Dynamics of HBV infection model with DNA-containing capsids, logistic hepatocyte growth and adaptive immune response,* Commun. Math. Biol. Neurosci. 2024 (2024), Article ID 58.

[15] Meskaf, A., Elkaf, M., and Allali, K. *Viral infection dynamics of HBV DNA-containing capsids with logistics growth and saturated rate,* Commun. Math. Biol. Neurosci., 2023 (2023), Article ID 7.

[16] Soriano, V., Moreno-Torres, V., Treviño, A., de Jesús, F., Corral, O., and de Mendoza, C. *Prospects for controlling hepatitis B globally.* Pathogens 13(4) (2024) 291.

[17] Sutradhar, R., and Dalal, D.C. *Re-cycling of DNA-containing capsids enhances hepatitis B virus infection,* arXiv preprint (2023) arXiv:2309.15665.

[18] Sutradhar, R., and Dalal, D.C. *Fractional-order models of hepatitis B virus infection with recycling effects of capsids,* Math. Methods Appl. Sci. 46(14) (2023) 15599–15625.

[19] World Health Organization. *Global health sector strategy on viral hepatitis 2016–2021: Towards ending viral hepatitis,* (2016).

[20] World Health Organization. *Advocacy brief: Strategic approaches to eliminate hepatitis B and C by 2030,* (2016).

Iranian Journal of Numerical Analysis and Optimization
Vol. 17, No. 1, 2027, pp 290–331
https://doi.org/10.22067/ijnao.2026.97941.1852
https://ijnao.um.ac.ir/

CC BY NC

Research Article

Memory-driven fractional dynamics of pneumonia transmission with prophylactic intervention

K. Soni, S. Kumar, S. Kumawat and A.K. Sinha*

Abstract

A novel fractional-order mathematical model is proposed to investigate the transmission dynamics of Pneumocystis pneumonia in an immunocompromised population. Long-term chemoprophylaxis is represented as an imperfect vaccination strategy within a susceptible-vaccinated-infected framework. The model is formulated using the Atangana–Baleanu fractional derivative in the Caputo sense to incorporate memory and hereditary effects in disease transmission. The positivity, boundedness, existence, uniqueness, and Ulam–Hyers stability of the proposed system are established. The basic reproduction number ($\mathfrak{R}_0$) is derived using the next-generation

*Corresponding author

Received 19 February 2026; revised 14 July 2026; accepted 25 July 2026

Kaushal Soni
Department of Mathematics, National Institute of Technology Raipur, India. e-mail: kaushalmath2205@gmail.com

Sanjeev Kumar
Department of Mathematics, Khandari Campus, Agra India. Maharaja Suhel Dev University Azamgarh India. e-mail: sanjeevibs@yahoo.co.in

Shyamsunder Kumawat
Department of Mathematics, SRM University Delhi-NCR, Sonepat, Haryana, India. e-mail: skumawatmath@gmail.com

Arvind Kumar Sinha
Department of Mathematics, National Institute of Technology Raipur, India. e-mail: aksinha.maths@nitrr.ac.in

How to cite this article
Soni, K., Kumar,S., Kumawat, S. and Sinha, A.K., Memory-driven fractional dynamics of pneumonia transmission with prophylactic intervention. *Iran. J. Numer. Anal. Optim.*, 2027; 17(1): 290-331. https://doi.org/10.22067/ijnao.2026.97941.1852

matrix approach, and the local asymptotic stability of the disease-free and endemic equilibrium points is investigated. A numerical scheme based on Lagrange interpolation is employed to obtain approximate solutions. Sensitivity analysis is performed to identify the parameters that most strongly influence disease transmission. The results indicate that increasing prophylaxis coverage and efficacy substantially reduces the basic reproduction number and suppresses disease spread. The proposed fractional-order framework provides a realistic description of pneumonia dynamics by incorporating memory effects and offers useful insights for evaluating prophylactic intervention strategies in vulnerable populations.

AMS subject classifications (2020): 26A33; 34A08; 34D20; 34K20; 37N25.

Keywords: Pneumonia transmission; Fractional-order model; Prophylaxis; Imperfect vaccine; Mathematical epidemiology; Stability analysis.

1 Introduction

Pneumonia is still one of the most serious health problems worldwide and is responsible for more than 700,000 children under 5 years dying each year, which is about one child every 43 seconds. It is also the top infectious cause of death among children globally [57]. The damage does not stop with children; in fact, last year, pneumonia led to more than one million deaths among people over 70 years of age, thus reflecting its heavy toll on people of all ages [51]. Pneumocystis pneumonia, caused by the fungal pathogen Pneumocystis jirovecii, is one of the serious complications of immunodeficiency and, therefore, a distinguishing infection with a growing threat. While originally it was one of the AIDS epidemic's hallmark diseases, Pneumocystis pneumonia has been increasingly detected in non-HIV immunocompromised patients. Thus, it can be very well seen in the cases of organ transplant recipients, people with cancer who undergo chemotherapy, and patients who are on long-term immunosuppressive treatment

[31, 27, 58, 12, 16]. The death toll of Pneumocystis pneumonia in non-HIV-infected individuals is alarmingly high, as it is reported to be between 30% and 50%, whereas in HIV-positive cases, it accounts for only $10 - 20\%$ [36, 19]. The major route of transmission is through the air, thereby spreading the disease from one person to another and leading to hospital-based outbreaks (nosocomial), notably in transplant units, where such outbreaks have been well documented. These outbreaks demonstrate the necessity of implementing infection prevention and control measures strictly [25, 60, 4]. Given its extremely high fatality rate and easy spread, clinical management is mainly characterised by the use of long-term chemoprophylaxis and early detection using advanced molecular diagnostic tools [50, 18, 8, 35]. Many theoretical approaches

have been developed to better understand the complex epidemiological dynamics of Pneumocystis pneumonia. Different mathematical modeling approaches have been employed to explore transmission mechanisms, study co-infection dynamics, evaluate the effects of interventions and fractional-order calculus [56, 55].

Mathematical models play a crucial role in analyzing how pneumonia spreads through populations and in developing interventions to reduce its prevalence. Several mathematical models employing integer-order derivatives have been constructed to analyze disease dynamics [45, 17, 13, 2]. Several key studies have focused on pneumonia transmission, including Manuputty, Lesnussa, and Leleury [26] witch investigated the stability and the impact of vaccination. Osiuigbo et al. [29] developed a model to study the dissemination of pneumonia. Kizito and Tumwiine [21] examined the combined impact of vaccination and treatment in controlling pneumococcal pneumonia. Kanyiri, Luboobi, and Kimathi [20] analyzed optimal strategies for controlling co-infection of influenza and pneumonia, taking into account the development of antiviral resistance. Teklu [54] examined the spread and transmission patterns of HIV/AIDS and pneumonia co-infection.

For a complex infection like Pneumocystis pneumonia, where transmission occurs within a specific, vulnerable subpopulation and preventative medication (prophylaxis) is a key intervention, mathematical modeling offers a unique opportunity to quantify the effects of prophylaxis coverage and efficacy on disease control. The application of mathematical modeling to pneumonia has yielded a growing body of literature aimed at understanding its transmission and control and several studies have employed integer-order differential equation models to explore intervention strategies.

Although studies [26, 29, 21, 20, 54] have significantly improved the understanding of pneumonia transmission, they mainly rely on classical integer-order differential equation models and investigate bacterial pneumonia, coinfection, vaccination, treatment, or optimal control strategies. None of these studies considers Pneumocystis pneumonia as a standalone disease in an immunocompromised population, nor do they model long-term chemoprophylaxis as an imperfect vaccination strategy. Furthermore, the existing models do not incorporate the Atangana–Baleanu fractional derivative with a nonsingular and nonlocal kernel, which enables the inclusion of memory and hereditary effects that are important for describing chronic disease progression and long-term prophylactic interventions. In addition, the proposed model establishes the positivity, boundedness, existence, uniqueness, Ulam–Hyers (UH) stability, equilibrium analysis, sensitivity analysis, and numerical simulations within a unified fractional-order framework. These features clearly distinguish the proposed model from the previously published studies. Therefore, the main novelty of the present work lies in the development of the first fractional-order susceptible-vaccinated-infected model for Pneumocystis pneumonia using the Atangana–Baleanu fractional derivative. Unlike previous studies, the proposed model treats chemoprophylaxis as an imperfect

vaccination strategy, explicitly incorporates memory effects through a nonsingular and nonlocal kernel, and provides a comprehensive theoretical and numerical analysis for assessing the impact of prophylaxis coverage and efficacy on disease transmission.

Several fractional operators utilizing power-law, exponential, generalized Mittag-Leffler (ML), and other kernel functions have been introduced, gaining notable interest in the modeling of biological processes [10, 11, 3]. Fractional-order operators offer advantages over classical-order operators, including memory effects and improved performance. Fractional-order operators have been successfully applied to Marburg virus, COVID-19, Hepatitis B, Zika virus, and Malaria disease modeling to investigate their spread and control [46, 5, 48, 47, 42, 37, 41, 49, 44, 38, 24, 43]. Specifically, Sabar et al. [33] examined strategies for managing pneumonia in livestock. Peter et al. [30] developed and studied a pneumococcal pneumonia model incorporating the Caputo–Fabrizio operator. They conducted numerical simulations and assessed the role of arbitrary fractional-order derivatives using model parameters, concluding that the force of infection plays a crucial role in controlling disease spread within a community.

Fractional-order differential equations offer a more precise modeling approach for real-world systems than conventional integer-order models [6, 59, 7, 23, 22]. Several fractional derivatives have been defined, broadening the classical integer-order derivative. Riemann and Liouville [34] laid the groundwork with their definition of the classical fractional-order derivative. One example of its application is the study of nematic liquid crystal diffusion using the fractional Hunter–Saxton equation [14]. Caputo and Fabrizio [11] introduced a fractional-order derivative characterized by an exponential kernel, further examined in [15, 32]. Subsequently, Atangana and Baleanu [3] proposed a fractional operator incorporating the ML function as its kernel, which possesses nonsingular and nonlocal properties. Unlike the Caputo–Fabrizio derivative, which uses a nonsingular exponential kernel, the Atangana–Baleanu fractional derivative in the Caputo sense employs a nonsingular and nonlocal Mittag-Leffler kernel. Hence, the Atangana–Baleanu fractional derivative in the Caputo sense is better suited to capture the long-term memory and hereditary effects involved in Pneumocystis pneumonia transmission.

While traditional integer-order epidemic models effectively describe local dynamics in the absence of external factors, they fail to capture the nonlocal interactions and crossover effects observed in many biological systems. Classical fractional-order operators, constrained by their singular kernels, also struggle to represent these complexities accurately. In contrast, the Atangana–Baleanu fractional derivative in the Caputo sense, with its generalized ML kernel, excels in modeling such intricate behaviors, including crossover effects in mean-square displacement. Additionally, its nonsingular kernel enhances the accuracy of predictions regarding both the early and late stages of disease transmission.

To date, no study has applied the sophisticated framework of fractional-order modeling to investigate the transmission dynamics of Pneumocystis Pneumonia as a standalone infection within its specific at-risk, immunocompromised population. Furthermore, the critical public health strategy of long-term chemoprophylaxis in this group has never been conceptually modeled as a form of "vaccination" within a fractional susceptible-vaccinated-infected framework with the Atangana–Baleanu fractional derivative in the Caputo sense. Therefore, our proposed study is uniquely positioned to bridge this gap by taking advantage of the methodological advantages of fractional calculus to analyze the specific epidemiology and the novel Pneumocystis pneumonia control paradigm.

Thus, to clearly address the identified gap in the present literature, this research project aims to develop and conduct a thorough analysis of a novel fractional-order susceptible-vaccinated-infected epidemiological model tailored to the transmission dynamics of Pneumocystis pneumonia. Firstly, one of the significant innovations of this article is to be among the handful of studies that have utilized a fractional-order modeling method in the particular epidemiological framework of Pneumocystis pneumonia. Through the Atangana–Baleanu fractional derivative in Caputo, the model can capture memory and hereditary features, which are essential for representing the nature of chronic infections and the susceptibility cycle of immunocompromised individuals over a long period. Secondly, the study, in a more conceptual and groundbreaking manner, redefines the long-term use of chemoprophylaxis, the main weapon against Pneumocystis pneumonia, as a form of imperfect vaccination within the model. This viewpoint not only provides a new theoretical basis for prophylactic assessment but also enhances understanding of its epidemiological effects at the population level.

This paper is structured in the following manner: Foundational principles of fractional calculus are presented in Section 2. Section 3 details the construction of the proposed mathematical model. The qualitative analysis of the model is the focus of Section 4, followed by an examination of UH stability in Section 5. In Section 6, we determine the basic reproduction number, equilibrium states, and stability analysis. A sensitivity analysis is conducted in Section 7. The numerical scheme implemented for simulations and the findings are discussed in Section 8. Finally, conclusions are drawn, and potential future research avenues are suggested in Section 9.

2 Preliminaries

This section begins with examining fundamental mathematical tools utilized in the following sections.

Definition 1 ([52]). Consider $m \in [1, \infty)$, and let $\mathbb{J}$ be an open subset. Then the Sobolev space $H^m(\mathbb{J})$ is expressed as follows:

$$H^m(\mathbb{J}) = \{\mathscr{U} \in L^2(\mathbb{J}) \big| \mathcal{D}^{\xi}\mathscr{U} \in L^2(\mathbb{J}) \ \forall |\xi| \leq m\}.$$

Atangana and Baleanu [3] formulated the Atangana–Baleanu fractional derivative in Caputo sense using this definition.

Definition 2 ([3]). Consider a function $\mathscr{G}(t) \in H^m(c, d)$, and let ξ be a fractional order such that $0 < \xi < 1$. Then, the Atangana–Baleanu fractional derivative in the Caputo sense is expressed as follows:

$$^{ABC}_{0}D_t^{\xi}\mathscr{G}(t) = \frac{ABC(\xi)}{(1-\xi)}\int_0^t E_{\xi}\left(\frac{-\xi}{1-\xi}(t-\varrho)^{\xi}\right)\mathscr{G}'d\varrho, \tag{1}$$

where $ABC(\xi)$ is a normalization mapping such that $ABC(0) = ABC(1) = 1$.

Definition 3 ([3]). The Laplace transform (LT) of the function $\mathscr{G}(t)$ is commonly defined as follows:

$$\mathscr{L}[\mathscr{G}(t), s] = \overline{\mathscr{G}}(s) = \int_0^{\infty} e^{-st}\mathscr{G}(t)dt, \qquad \mathbb{R}(s) > \zeta, \quad t \geq 0. \tag{2}$$

Taking inverse LT of $\overline{\mathscr{G}}(s)$, we have

$$\mathscr{L}^{-1}[\overline{\mathscr{G}}(s), t] = \frac{1}{2\pi i}\int_{\eta-i\infty}^{\eta+i\infty} e^{st}\overline{\mathscr{G}}(s)ds, \text{where} \ \ \eta \in \mathbb{R}. \tag{3}$$

Definition 4 ([3]). The LT of Atangana–Baleanu fractional derivative in the Caputo sense is expressed as follows:

$$\mathscr{L}\left[^{ABC}_{0}D_t^{\xi}\mathscr{G}(t)\right] = \frac{ABC(\xi)}{1-\xi}\left[\frac{s^{\xi}\mathscr{L}\{\mathscr{G}(t)\} - s^{\xi-1}\mathscr{G}(0)}{s^{\xi} + \frac{\xi}{1-\xi}}\right]. \tag{4}$$

Definition 5 ([3]). Let $\mathscr{G}(t) \in H^m(a, b)$, and consider a fractional order ξ with $0 < \xi < 1$ then the fractional integral in the Atangana–Baleanu sense is defined as follows:

$$^{ABC}_{0}\mathfrak{I}_t^{\xi}\mathscr{G}(t) = \frac{1-\xi}{ABC(\xi)}\mathscr{G}(t) + \frac{\xi}{ABC(\xi)\Gamma(\xi)}\int_0^t (t-\tau)^{\xi-1}\mathscr{G}(\tau)d\tau. \tag{5}$$

Definition 6. For $\Psi > 0$, the Mittag-Leffler function $E_\Psi(\mathfrak{Z})$ was first proposed by Mittag-Leffler, while its two-parameter generalization, $E_{\Psi,\varpi}(\mathfrak{Z})$, is credited to Wiman,

$$E_\Psi(\mathfrak{Z}) = \sum_{n=0}^{\infty} \frac{\mathfrak{Z}^k}{\Gamma(\Psi k + 1)} \qquad (\mathfrak{Z}, \Psi \in \mathbb{C}, \mathbb{R}(\Psi) > 0), \tag{6}$$

$$E_{\Psi,\varpi}(\mathfrak{Z}) = \sum_{n=0}^{\infty} \frac{\mathfrak{Z}^k}{\Gamma(\Psi k + \varpi)}, \qquad (\mathfrak{Z}, \Psi, \varpi \in \mathbb{C}, \mathbb{R}(\Psi) > 0, \mathbb{R}(\varpi) > 0). \tag{7}$$

Lemma 1 ([39])**.** For $0 < \xi \leq 1$, we consider the fractional differential problem described as follows:

$${}^{ABC}_{0}D_t^{\xi}\mathscr{G}(t) = \mathscr{B}(t), \qquad t \in [0,T], \quad \mathscr{G}(0) = \mathscr{G}_0, \tag{8}$$

where $\mathscr{G} : [0,T] \to \mathbb{R}$ is a function and $\mathscr{G}_0$ is a fixed point such that

$$\mathscr{G}(t) = \mathscr{G}_0 + \frac{1-\xi}{ABC(\xi)}\mathscr{B}(t) + \frac{\xi}{ABC(\xi)\Gamma(\xi)}\int_0^t (t-\tau)^{\xi-1}\mathscr{B}(\tau)d\tau. \tag{9}$$

Lemma 2. Let $\phi : D \to \mathbb{R}^n$ with $\phi \in C^1(D)$ and let $x \in D \subseteq \mathbb{R}$. Then ϕ fulfills a Lipschitz condition on each convex compact set $D \subseteq D$ with Lipschitz constant $\mathbb{K} > 0$ such that

$$\mathbb{K} = \max_{x \in D}\left|\frac{d\phi}{dx}\right|.$$

Lemma 3. Let $\mathcal{F}(t) \in C[a,b]$, and assume its Atangana–Baleanu fractional derivative ${}^{ABC}_{0}D_t^{\xi}\mathcal{F}(t) \in C[a,b]$ for $0 < \xi \leq 1$. Then, for every $t \in [a,b]$ there exists a point $l_1 \in [a,t]$ such that

$$\mathcal{F}(t) - \mathcal{F}(a) = \left[1 - E_\xi\left(-\frac{\xi(t-c_1)^\xi}{1-\xi}\right)\right]^{-1} \frac{(t-a)^\xi}{ABC(\xi)\Gamma(\xi)}\,{}^{ABC}_{0}D_t^{\xi}\mathcal{F}(l_1), \tag{10}$$

where $c_1 \in (a,t)$.

Lemma 4 (Generalized mean value theorem [28])**.** Let $g(t) \in C[a,b]$, and let ${}^{ABC}_{0}D_t^{\xi}g(t) \in C[a,b]$ when $0 < \xi \leq 1$ Then $g(t) = g(a) + \frac{1}{\Gamma(\xi)}\,{}^{ABC}_{0}D_t^{\xi}g(\delta)(t-a)^\xi$, when $0 \leq \delta \leq t$, for all $t \in (a,b]$

Note that by Lemma 4 if $g(t) \in (0,b]$, and ${}^{ABC}_{0}D^{\xi}_{t}g(t) \in C(a,b]$ ${}^{ABC}_{0}D^{\xi}_{t}g(t) \leq 0$ for every $t \in (0,b]$, the function $g(t)$ is nonincreasing on for all $t \in (0,b]$ and ${}^{ABC}_{0}D^{\xi}_{t}gF(t) \geq 0$ for every $t \in (0,b]$, then the function $g(t)$ is nondecreasing for all t in $(0,b]$.

Theorem 1 (Schaefer's fixed point theorem [9])**.** Let V be a normed linear space, and let T be a continuous mapping of V into V that is compact on each bounded subset X of V. Then set $W = \{x \in V : \ x = \lambda T(x) \text{ for some } 0 \leq \lambda \leq 1\}$ is bounded. Then T has a fixed point.

3 Model formulation

To investigate the transmission dynamics of Pneumocystis Pneumonia, a serious fungal infection, we propose a compartmental model based on its modern epidemiological understanding. Caused by the fungus Pneumocystis jirovecii, Pneumocystis Pneumonia is now understood to spread directly from person to person via an airborne route [25]. The disease primarily poses a threat to individuals with compromised immune systems, such as organ transplant recipients, patients undergoing chemotherapy, or those with HIV/AIDS [53, 27]. A key strategy in managing Pneumocystis Pneumonia is the use of long-term chemoprophylaxis (preventative medication) in these at-risk populations [50]. In this study, we innovatively model this public health intervention as a form of vaccination. This allows us to use a susceptible-vaccinated-infected ($\mathcal{XYZ}$) framework to analyze the impact of prophylaxis coverage and efficacy on the spread of the disease. The model is specifically designed to capture the dynamics within the vulnerable, immunocompromised population, where the infection is most prevalent and severe. In this work, the total population $\mathcal{N}$ is partitioned into three mutually exclusive compartments. The class $\mathcal{X}$ denotes immunocompromised individuals who are susceptible to Pneumocystis Pneumonia and are not receiving prophylactic protection. The class $\mathcal{Y}$ represents the population of immunocompromised individuals who are receiving prophylactic treatment. This group is analogous to a vaccinated population, having a significantly reduced, but not zero, probability of infection. The class $\mathcal{Z}$ represents the population of individuals who are actively infected with Pneumocystis Pneumonia and are capable of transmitting the fungus to individuals in both the $\mathcal{X}$ and $\mathcal{Y}$ compartments.

To account for the memory and nonlocal effects associated with the progression and transmission of Pneumocystis Pneumonia, the proposed $\mathcal{XYZ}$ model is formulated using the Atangana–Baleanu fractional derivative in Caputo. The Atangana–Baleanu fractional derivative in Caputo is characterized by a nonsingular kernel, which allows past states of the system to influence the current dynamics in a biologically realistic manner. The governing equations are considered in the Atangana–Baleanu fractional derivative in Caputo sense with order $0 < \xi \leq 1$ as follows:

$$
\begin{aligned}
{}^{ABC}_{0}D_t^{\xi}\mathcal{X}(t) &= (1-p)\Lambda - \frac{\beta_1 \mathcal{X}\mathcal{Z}}{\mathcal{X}+\mathcal{Y}+\mathcal{Z}} - \mu_1 \mathcal{X},\\
{}^{ABC}_{0}D_t^{\xi}\mathcal{Y}(t) &= p\Lambda - \frac{(1-\psi)\beta_2 \mathcal{Y}\mathcal{Z}}{\mathcal{X}+\mathcal{Y}+\mathcal{Z}} - \mu_2 \mathcal{Y},\\
{}^{ABC}_{0}D_t^{\xi}\mathcal{Z}(t) &= \frac{\beta_1 \mathcal{X}\mathcal{Z}}{\mathcal{X}+\mathcal{Y}+\mathcal{Z}} + \frac{(1-\psi)\beta_2 \mathcal{Y}\mathcal{Z}}{\mathcal{X}+\mathcal{Y}+\mathcal{Z}} - \mu_3 \mathcal{Z},
\end{aligned} \tag{11}
$$

with initial conditions $\mathfrak{P}_0 = (\mathcal{X}_0, \mathcal{Y}_0, \mathcal{Z}_0)$ and total population size $\mathcal{N} = \mathcal{X} + \mathcal{Y} + \mathcal{Z}$.

The formulation of our model rests on the following key assumptions about Pneumocystis Pneumonia. Transmission mechanism: Infection occurs through airborne exposure, whereby individuals in the infected class ($\mathcal{Z}$) transmit the fungus to both susceptible ($\mathcal{X}$) and prophylaxis-protected ($\mathcal{Y}$) individuals. The resulting infection pressure is described using a standard incidence function. Target population: The model focuses solely on the vulnerable, immunocompromised population. The recruitment parameter Λ denotes the rate at which individuals newly enter this at-risk group, such as through organ transplantation, initiation of immunosuppressive treatment, or new HIV diagnoses. Prophylactic intervention: Upon entry into the at-risk population, a proportion p of individuals is assumed to immediately initiate chemoprophylaxis and therefore join the protected class $\mathcal{Y}$, while the remaining fraction $(1-p)$ enters the susceptible class $\mathcal{X}$ Imperfect protection: Prophylactic treatment is assumed to provide partial protection. With efficacy level ψ, individuals in the protected class $\mathcal{Y}$ remain susceptible to infection, though their risk is reduced by a factor of $(1-\psi)$.

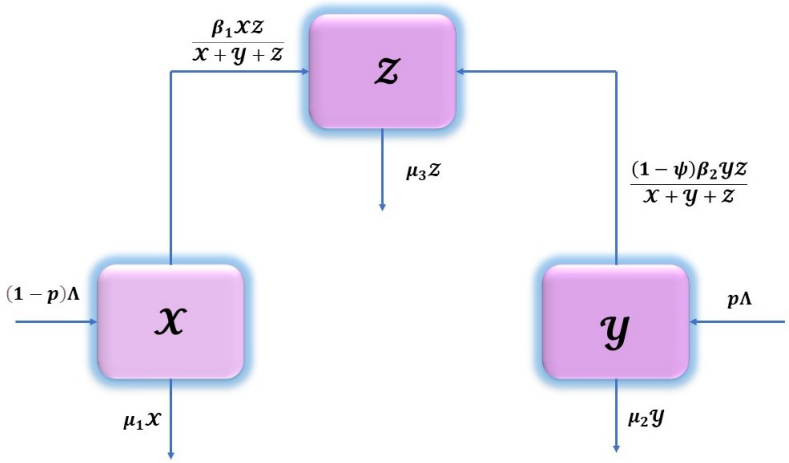


Figure 1: Schematic diagram for the Pneumonia transmission.

According to the model description and the parameters in Table 1, the schematic diagram is represented by Figure 1 below.

Table 1: Description of parameters in model (11).

Parameters	Description	value	Source
Λ	Recruitment rate of susceptible immunocompromised individuals	10.9	[21]
β_1	Transmission rate from infected to unprotected susceptibles	0.0287	[21]
β_2	Transmission rate from infected to prophylaxis-protected	0.0286	Assumed
μ_1	Natural mortality rate of susceptible individuals without protection	0.0002	[1]
μ_2	Natural Mortality rate of prophylaxis-protected individuals	0.0001	Fitted
μ_3	Total Mortality rate of infected individuals	0.3302	[1]
ψ	Efficacy of the prophylactic drug regimen	[0,1]	Assumed
p	Coverage rate of prophylaxis among new at-risk individuals	[0,1]	Assumed

4 Qualitative analysis

This section provides a mathematical analysis of solution behavior, such as positivity, existence, uniqueness, and boundedness.

4.1 Positivity and boundedness

This part focuses on proving that,given all initial state variables are positive,the model consistently yields solutions that are not only finite but also stay nonnegative for every $t > 0$. The proof relies on the application of the subsequent theorem.

Theorem 2. The proposed infection system (11) provides positive and bounded solutions for positive initial conditions.

Proof. We first prove that every solution $(\mathcal{X}(t), \mathcal{Y}(t), \mathcal{Z}(t))$ of system (11) is nonnegative for all $t > 0$; that is, $\mathcal{X}(t), \mathcal{Y}(t)$, and $\mathcal{Z}(t)$ remain greater than or equal to zero.
To show that the set

$$\Omega = \left\{ (\mathcal{X}, \mathcal{Y}, \mathcal{Z}) : 0 \leq \mathcal{X} + \mathcal{Y} + \mathcal{Z} \leq \frac{\Lambda}{\mu_1 + \mu_2 + \mu_3} \right\}. \tag{12}$$

is positively invariant, using Lemma 4

$$\begin{cases} {}^{ABC}_{0}D_t^{\xi}\mathcal{X}(t)\Big|_{\mathcal{X}(t')=0} = (1-p)\Lambda \geq 0, \\ {}^{ABC}_{0}D_t^{\xi}\mathcal{Y}(t)\Big|_{\mathcal{Y}(t')=0} = p\Lambda \geq 0, \\ {}^{ABC}_{0}D_t^{\xi}\mathcal{Z}(t)\Big|_{\mathcal{Z}(t')=0} = 0 \geq 0. \end{cases} \tag{13}$$

It follows from (13) that each of the solutions of (11) is nonnegative and remains in $\mathbb{R}^3_+$, and hence the set Ω defined in (12) is positively invariant for model (11). Next, we introduce the total population $\mathcal{N}(t) = \mathcal{X}(t) + \mathcal{Y}(t) + \mathcal{Z}(t)$, then

$$\begin{aligned} {}^{ABC}_{0}D_t^{\xi}\mathcal{N}(t) &= {}^{ABC}_{0}D_t^{\xi}\mathcal{X}(t) + {}^{ABC}_{0}D_t^{\xi}\mathcal{Y}(t) + {}^{ABC}_{0}D_t^{\xi}\mathcal{Z}(t), \\ {}^{ABC}_{0}D_t^{\xi}\mathcal{N}(t) &= \Lambda - \mu_1\mathcal{X}(t) + \mu_2\mathcal{Y}(t) + \mu_3\mathcal{Z}(t), \\ {}^{ABC}_{0}D_t^{\xi}\mathcal{N}(t) &\leq \Lambda - (\mu_1+\mu_2+\mu_3)\mathcal{N}(t), \\ \mathcal{N}(t) &\leq \left(\frac{ABC(\xi)}{ABC(\xi)+(1-\xi)(\mu_1+\mu_2+\mu_3)}\mathcal{N}(0) \right. \\ &\quad \left. + \frac{(1-\xi)\Lambda}{ABC(\xi)+(1-\xi)(\mu_1+\mu_2+\mu_3)}\right) \\ &\quad E_{\xi,1}\left(-\frac{\xi(\mu_1+\mu_2+\mu_3)}{ABC(\xi)+(1-\xi)(\mu_1+\mu_2+\mu_3)}t^{\xi}\right) \\ &\quad + \frac{\xi\Lambda}{ABC(\xi)+(1-\xi)(\mu_1+\mu_2+\mu_3)}E_{\xi,\xi+1} \\ &\quad \left(-\frac{\xi(\mu_1+\mu_2+\mu_3)}{ABC(\xi)+(1-\xi)(\mu_1+\mu_2+\mu_3)}t^{\xi}\right) \\ &\to \frac{\Lambda}{\mu_1+\mu_2+\mu_3} \text{ as } t\to\infty. \end{aligned}$$

Hence (12) is the biologically feasible region of model (11). □

4.2 Existence results

This section focuses on analyzing the qualitative dynamics of the system (11). We begin by introducing

$$\begin{cases} {}^{ABC}_{0}D_t^{\xi}\mathcal{X}(t) = \aleph_1(t,\mathcal{X},\mathcal{Y},\mathcal{Z}), \\ {}^{ABC}_{0}D_t^{\xi}\mathcal{Y}(t) = \aleph_2(t,\mathcal{X},\mathcal{Y},\mathcal{Z}), \\ {}^{ABC}_{0}D_t^{\xi}\mathcal{Z}(t) = \aleph_3(t,\mathcal{X},\mathcal{Y},\mathcal{Z}), \end{cases} \tag{14}$$

where

$$\begin{cases} \aleph_1(t,\mathcal{X},\mathcal{Y},\mathcal{Z}) = (1-p)\Lambda - \frac{\beta_1\mathcal{X}\mathcal{Z}}{\mathcal{X}+\mathcal{Y}+\mathcal{Z}} - \mu_1\mathcal{X}, \\ \aleph_2(t,\mathcal{X},\mathcal{Y},\mathcal{Z}) = p\Lambda - \frac{(1-\psi)\beta_2\mathcal{Y}\mathcal{Z}}{\mathcal{X}+\mathcal{Y}+\mathcal{Z}} - \mu_2\mathcal{Y}, \\ \aleph_3(t,\mathcal{X},\mathcal{Y},\mathcal{Z}) = \frac{\beta_1\mathcal{X}\mathcal{Z}}{\mathcal{X}+\mathcal{Y}+\mathcal{Z}} + \frac{(1-\psi)\beta_2\mathcal{Y}\mathcal{Z}}{\mathcal{X}+\mathcal{Y}+\mathcal{Z}} - \mu_3\mathcal{Z}, \end{cases} \tag{15}$$

which can be expressed as

$$\begin{cases} {}^{ABC}_{0}D_t^{\xi}\aleph(t) = \eth(t,\aleph(t)), & t \in [0,T], \\ \aleph(0) = \aleph_0, & 0 < \xi \leq 1, \end{cases} \tag{16}$$

where $\aleph(t) = (\mathcal{X}(t),\mathcal{Y}(t),\mathcal{Z}(t)^T$ and $\aleph_0 = (\mathcal{X}_0,\mathcal{Y}_0,\mathcal{Z}_0)^T$.

$${}^{ABC}_{0}D_t^{\xi}\aleph(t) = \begin{cases} {}^{ABC}_{0}D_t^{\xi}\mathcal{X}(t), \\ {}^{ABC}_{0}D_t^{\xi}\mathcal{Y}(t), \\ {}^{ABC}_{0}D_t^{\xi}\mathcal{Z}(t). \end{cases} \qquad \eth(t,\aleph(t)) = \begin{cases} \aleph_1(t,\mathcal{X},\mathcal{Y},\mathcal{Z}), \\ \aleph_2(t,\mathcal{X},\mathcal{Y},\mathcal{Z}), \\ \aleph_3(t,\mathcal{X},\mathcal{Y},\mathcal{Z}). \end{cases} \tag{17}$$

Let $\mathbb{P} = C([0,T],\mathbb{R})$ signify the Banach space consisting of all continuous functions from the interval $[0,T]$ to $\mathbb{R}$, endowed with the norm given as

$||\aleph|| = \max_{t\in[0,T]} |\aleph(t)|$, where $|\aleph(t)| = |\mathcal{X}(t)| + \mathcal{Y}(t)| + |\mathcal{Z}(t)|$.

Based on Lemma 1, the solution of system (16) can be expressed as follows:

$$\aleph(t) = \aleph_0 + \frac{1-\xi}{ABC(\xi)}\eth(t,\aleph(t)) + \frac{\xi}{ABC(\xi)\Gamma(\xi)}\int_0^t (t-\tau)^{\xi-1}\eth(\tau,\aleph(\tau))d\tau. \tag{18}$$

Lemma 5. The following conditions are key to our analysis:

(I) There exist two positive constants $\mathfrak{W}_1$, $\mathfrak{W}_2$ such that

$$|\eth(t,\aleph(t))| \leq \mathfrak{W}_1||\aleph|| + \mathfrak{W}_2. \tag{19}$$

(II) There exists $\mathfrak{W}_3 > 0$ such that

$$|\eth(t, \aleph(t)) - \eth(t, \overline{\aleph}(t))| \leq \mathfrak{W}_3 |\aleph(t) - \overline{\aleph}(t)|. \tag{20}$$

Proof. From system (15), we have

$$\eth(t, \aleph(t)) = \begin{pmatrix} (1-p)\Lambda - \frac{\beta_1 \mathcal{X}\mathcal{Z}}{\mathcal{X}+\mathcal{Y}+\mathcal{Z}} - \mu_1 \mathcal{X} \\ p\Lambda - \frac{(1-\psi)\beta_2 \mathcal{Y}\mathcal{Z}}{\mathcal{X}+\mathcal{Y}+\mathcal{Z}} - \mu_2 \mathcal{Y} \\ \frac{\beta_1 \mathcal{X}\mathcal{Z}}{\mathcal{X}+\mathcal{Y}+\mathcal{Z}} + \frac{(1-\psi)\beta_2 \mathcal{Y}\mathcal{Z}}{\mathcal{X}+\mathcal{Y}+\mathcal{Z}} - \mu_3 \mathcal{Z} \end{pmatrix}. \tag{21}$$

Using the triangle inequality, we obtain

$$\begin{aligned} \left|\eth(t, \aleph(t))\right| \leq & \left|(1-p)\Lambda - \frac{\beta_1 \mathcal{X}\mathcal{Z}}{\mathcal{X}+\mathcal{Y}+\mathcal{Z}} - \mu_1 \mathcal{X}\right| \\ & + \left|p\Lambda - \frac{(1-\psi)\beta_2 \mathcal{Y}\mathcal{Z}}{\mathcal{X}+\mathcal{Y}+\mathcal{Z}} - \mu_2 \mathcal{Y}\right| \\ & + \left|\frac{\beta_1 \mathcal{X}\mathcal{Z}}{\mathcal{X}+\mathcal{Y}+\mathcal{Z}} + \frac{(1-\psi)\beta_2 \mathcal{Y}\mathcal{Z}}{\mathcal{X}+\mathcal{Y}+\mathcal{Z}} - \mu_3 \mathcal{Z}\right|. \end{aligned}$$

Again applying the triangle inequality gives

$$\begin{aligned} \left|\eth(t, \aleph(t))\right| \leq & \left|(1-p)\Lambda\right| + \beta_1 \mathbb{M}_1 + \mu_1 \left|\mathcal{X}\right| \\ & + \left|p\Lambda\right| + (1-\psi)\beta_2 \mathbb{M}_2 + \mu_2 \left|\mathcal{Y}\right| \\ & + \beta_1 \mathbb{M}_1 + (1-\psi)\beta_2 \mathbb{M}_2 + \mu_3 \left|\mathcal{Z}\right|, \end{aligned}$$

where $\mathbb{M}_1 = \max\limits_{t\in[0,T]} \left|\frac{\mathcal{X}\mathcal{Z}}{\mathcal{X}+\mathcal{Y}+\mathcal{Z}}\right|$ and $\mathbb{M}_2 = \max\limits_{t\in[0,T]} \left|\frac{\mathcal{Y}\mathcal{Z}}{\mathcal{X}+\mathcal{Y}+\mathcal{Z}}\right|$. □

Therefore,

$$\left|\eth(t, \aleph(t))\right| \leq \underbrace{\Lambda + 2\beta_1 \mathbb{M}_1 + 2(1-\psi)\beta_2 \mathbb{M}_2}_{\mathfrak{M}_2} + \underbrace{(\mu_1 + \mu_2 + \mu_3)}_{\mathfrak{M}_1} ||\aleph||,$$

$$\left|\eth(t, \aleph(t))\right| \leq \mathfrak{M}_1 ||\aleph|| + \mathfrak{M}_2.$$

Next, we prove condition (II). Let $\mathfrak{G} = \left\{(t, \aleph) | 0 \leq t \leq T, \aleph \in \mathbb{R}^3_+\right\}$ and let $\mathscr{U} \subseteq \mathfrak{G}$ be convex and compact. Let $\aleph, \overline{\aleph} \in \mathfrak{G}$. Then by the mean value theorem, there exists $\theta \in (\aleph, \overline{\aleph})$ such that

$$\begin{aligned}\frac{\eth(\aleph(t))-\eth(\overline{\aleph}(t))}{\aleph(t)-\overline{\aleph}(t)}&=\eth'(\theta(t)),\\ \eth(\aleph(t))-\eth(\overline{\aleph}(t))&=\eth'(\theta(t))(\aleph(t)-\overline{\aleph}(t)),\\ \left|\eth(\aleph(t))-\eth(\overline{\aleph}(t))\right|&=\left|\eth'(\theta(t))(\aleph(t)-\overline{\aleph}(t))\right|\\ &\leq \left|\left|\eth'(\theta)\right|\right|_{\infty}||\aleph-\overline{\aleph}||.\end{aligned}$$

Since $\eth\in C^1[0,T]$, over convex compact set $\mathscr{U}$, there exists a constant $\mathfrak{M}_3>0$ such that

$$\left|\left|\eth'(\theta)\right|\right|_{\infty}\leq\mathfrak{M}_3,$$

hence,

$$\begin{aligned}\left|\eth(\aleph(t))-\eth(\overline{\aleph}(t))\right|&\leq\mathfrak{M}_3\left|\left|\aleph-\overline{\aleph}\right|\right|_{\infty},\\ \max_{t\in[0,T]}\left|\eth(\aleph)-\eth(\overline{\aleph})\right|&\leq\mathfrak{M}_3\left|\left|\aleph-\overline{\aleph}\right|\right|_{\infty},\\ \left|\left|\eth(\aleph)-\eth(\overline{\aleph})\right|\right|_{\infty}&\leq\mathfrak{M}_3\left|\left|\aleph-\overline{\aleph}\right|\right|_{\infty}.\end{aligned}$$

Hence $\eth(\aleph)$ is Lipschitz.

We start with the following result.

Lemma 6. Let us consider a map $\widehat{\mathscr{Q}}$ on $\mathbb{P}$ such that

$$\widehat{\mathscr{Q}}(\aleph(t))=\aleph_0+\tfrac{1-\xi}{ABC(\xi)}\eth(t,\aleph(t))+\tfrac{\xi}{ABC(\xi)\Gamma(\xi)}\int_0^t(t-\tau)^{\xi-1}\eth(\tau,\aleph(\tau))d\tau. \tag{22}$$

Then it is completely continuous.

Proof. Initially, we observe that if $\widehat{\mathscr{Q}}\in\mathbb{P}$, then $\widehat{\mathscr{Q}}(\aleph)\in\mathbb{P}$. For all $\aleph\in\mathscr{M}_z=\{\aleph\in\mathbb{P}\ni||\aleph||\leq z\}$, we have

$$\begin{aligned}|\widehat{\mathscr{Q}}(\aleph(t))|\leq&||\aleph_0||+\frac{1-\xi}{ABC(\xi)}\max_{t\in[0,T]}|\eth(t,\aleph(t))|\\ &+\max_{t\in[0,T]}|\frac{\xi}{ABC(\xi)\Gamma(\xi)}\times\int_0^t(t-\tau)^{\xi-1}\eth(\tau,\aleph(\tau))d\tau|\\ \leq&||\aleph_0||+\frac{1-\xi}{ABC(\xi)}(\mathfrak{W}_1||\aleph||+\mathfrak{W}_2)+\frac{T^{\xi}}{ABC(\xi)\Gamma(\xi)}(\mathfrak{W}_1||\aleph||+\mathfrak{W}_2)\\ \leq&||\aleph_0||+\left(\frac{1-\xi}{ABC(\xi)}+\frac{T^{\xi}}{ABC(\xi)\Gamma(\xi)}\right)\mathfrak{W}_2\end{aligned}$$

$$+\left(\frac{1-\xi}{ABC(\xi)}+\frac{T^{\xi}}{ABC(\xi)\Gamma(\xi)}\right)\mathfrak{W}_1||\aleph||$$
$$\leq||\aleph_0||+\daleth(\mathfrak{W}_1||\aleph||+\mathfrak{W}_2)<\infty, \tag{23}$$

where $\daleth=\left(\frac{1-\xi}{ABC(\xi)}+\frac{T^{\xi}}{ABC(\xi)\Gamma(\xi)}\right)$. This implies that $||\widehat{\mathscr{Q}}(\aleph)||<\infty$. Therefore, $\widehat{\mathscr{Q}}$ is uniformly bounded on $\mathscr{M}_z$.

Next, we establish that the operator $\widehat{\mathscr{Q}}$ is equicontinuous. To achieve this, let $\aleph\in\mathscr{M}_z$ and let $t_1,t_2\in[0,T]$ such that $t_1<t_2$. Then

$$\begin{aligned}|\widehat{\mathscr{Q}}\aleph(t_1)-\widehat{\mathscr{Q}}\aleph(t_2)|&\leq\max_{t\in[0,T]}\left|\frac{1-\xi}{ABC(\xi)}(\eth(t_1,\aleph(t_1))-\eth(t_2,\aleph(t_2))\right|\\&\quad+\max_{t\in[0,T]}\frac{\xi}{ABC(\xi)\Gamma(\xi)}\left|\int_0^{t_1}(t_1-\tau)^{\xi-1}\eth(\tau,\aleph(\tau))d\tau\right.\\&\quad\left.-\int_0^{t_2}(t_2-\tau)^{\xi-1}\eth(\tau,\aleph(\tau))d\tau\right|\\&\leq\frac{1-\xi}{ABC(\xi)}\max_{t\in[0,T]}\left|\eth(t_1,\aleph(t_1))-\eth(t_2,\aleph(t_2)\right|\\&\quad+\frac{\mathfrak{W}_1||\aleph||+\mathfrak{W}_2}{ABC(\xi)\Gamma(\xi)}(t_2^{\xi}-t_1^{\xi}).\end{aligned} \tag{24}$$

It is evident that the right-hand side of the above inequality approaches zero as $t_1\to t_2$. The operator $\widehat{\mathscr{Q}}$ is equicontinuous, and from (23), the operator $\widehat{\mathscr{Q}}$ is uniformly bounded. Therefore, by the Arzela-Ascoli theorem, it follows that $\widehat{\mathscr{Q}}$ is relatively compact, implying that $\widehat{\mathscr{Q}}$ is completely continuous. □

Theorem 3. Model (11) has at least one solution.

Proof. Consider $\widehat{\Omega}$ such that

$$\widehat{\Omega}=\left\{\aleph\in\mathbb{P}\ni\aleph=\upsilon\widehat{\mathscr{Q}}(\aleph),\upsilon\in(0,1)\right\}.$$

In order to establish that the set $\widehat{\Omega}$ is bounded, suppose that $\aleph\in\mathscr{M}_z$, and confirm that the following inequality is satisfied for all $t\in[0,T]$:

$$||\aleph||=\upsilon||\widehat{\mathscr{Q}}(\aleph)||\leq\upsilon\left(|\aleph_0|+\left(\frac{1-\xi}{ABC(\xi)}+\frac{T^{\xi}}{ABC(\xi)\Gamma(\xi)}\right)\left[\mathfrak{W}_1||\aleph||+\mathfrak{W}_2\right]\right). \tag{25}$$

Hence, the set $\widehat{\Omega}$ is proven to be bounded. Therefore, by invoking Schaefer's fixed point theorem, we deduce that the operator $\widehat{\mathscr{Q}}(\aleph)$ possesses at least one fixed point. Consequently, the proposed pneumonia transmission model (11) admits at least one solution.

□

Theorem 4. The solution to (11) is guaranteed to be unique provided the following condition is satisfied:

$$\tfrac{1-\xi}{ABC(\xi)}\mathfrak{W}_3 + \tfrac{T^{\xi}}{ABC(\xi)\Gamma(\xi)}\mathfrak{W}_3 < 1. \tag{26}$$

Proof. To obtain the required result, we apply the Banach contraction theorem. Let $\aleph$, $\overline{\aleph} \in \mathbb{P}$, and assume that condition (II) holds. Then,

$$\begin{aligned}
||\widehat{\mathscr{Q}}\aleph - \widehat{\mathscr{Q}}\overline{\aleph}|| \leq & \max_{t\in[0,T]} |\frac{1-\xi}{ABC(\xi)}\eth(t,\aleph(t)) - \frac{1-\xi}{ABC(\xi)}\eth(t,\overline{\aleph}(t)) \\
& + \frac{\xi}{ABC(\xi)\Gamma(\xi)}\int_0^t (t-\tau)^{\xi-1}\eth(\tau,\aleph(\tau))d\tau \\
& + \frac{\xi}{ABC(\xi)\Gamma(\xi)}\int_0^t (t-\tau)^{\xi-1}\eth(\tau,\overline{\aleph}(\tau))d\tau| \\
\leq & \left(\frac{1-\xi}{ABC(\xi)}\mathfrak{W}_3 + \frac{T^{\xi}}{ABC(\xi)\mathfrak{W}_3}\right)||\aleph - \overline{\aleph}||.
\end{aligned} \tag{27}$$

As a result, we derive

$$||\widehat{\mathscr{Q}}\aleph - \widehat{\mathscr{Q}}\overline{\aleph}|| \leq \mathscr{B}||\aleph - \overline{\aleph}||, \tag{28}$$

where

$$\mathscr{B} = \tfrac{1-\xi}{ABC(\xi)}\mathfrak{W}_3 + \tfrac{T^{\xi}}{ABC(\xi)\Gamma(\xi)}\mathfrak{W}_3.$$

Therefore, we conclude that there exists a unique fixed point of $\widehat{\mathscr{Q}}$, which implies that system (11) admits a unique solution. □

5 UH stability

UH stability is crucial in ensuring that the solutions of fractional differential equations (FDEs) are resistant to small perturbations, which is essential when modeling systems with memory effects and nonlocal dynamics. This stability investigation helps determine whether small changes in the initial data or parameters lead to limited variations in the solutions, which is essential for verifying the reliability and strength of the proposed model. Recent research has applied UH stability to fractional functional differential equations and neutral fractional evolution equations with delay, demonstrating its significance in preserving the reliability and practical utility of fractional models in diverse fields, including biology and engineering.

We examine the stability of the proposed system by introducing a small perturbation $\mathfrak{F} \in C[0,T]$, which helps assess the system's response to minor changes. This analysis is essential for confirming the system's resilience under different conditions:

(a) $||\mathfrak{F}(t)|| \leq \varepsilon,\ \mathfrak{F} = \max(\mathfrak{F}_j)^T,\ t \in [0,T],\ j = 1,2,\ldots,5,$

(b) ${}^{ABC}_{0}D^{\xi}_{t}\aleph(t) = \eth(t,\aleph(t)) + \mathfrak{F}(t),\ t \in [0,T].$

Lemma 7. The solution of the modified problem can be expressed as follows:

$$ {}^{ABC}_{0}D^{\xi}_{t}\aleph(t) = \eth(t,\aleph(t)) + \mathfrak{F}(t), \qquad t \in [0,T], \tag{29}$$

with the initial condition

$\aleph(0) = \aleph_0,$

with the relationship

$$\begin{aligned} &\left|\left|\aleph(t) - \left(\aleph_0(t) + \tfrac{1-\xi}{ABC(\xi)}\eth(t,\aleph(t)) \right.\right.\right. \\ &\left.\left.\left. + \frac{\xi}{ABC(\xi)\Gamma(\xi)}\int_0^t (t-\tau)^{\xi-1}\eth(\tau,\aleph(\tau))d\tau\right)\right|\right| \leq \epsilon\mathfrak{U}, \end{aligned} \tag{30}$$

where $\mathfrak{U} = \frac{(\Gamma(\xi)(1-\xi)+T^{\xi})}{ABC(\xi)\Gamma(\xi)}$ holds true.

Proof. From (30), we have ${}^{ABC}_{0}D^{\xi}_{t}\aleph(t) = \eth(t,\aleph(t)) + \mathfrak{F}(t),\ t \in [0,T]$. Now, we use Definition 5 and obtain

$$\left.\begin{aligned} \aleph(t) = \aleph_0(t) + \tfrac{1-\xi}{ABC(\xi)}\eth(t,\aleph(t)) + \frac{\xi}{ABC(\xi)\Gamma(\xi)}\int_0^t (t-\tau)^{\xi-1}\eth(\tau,\aleph(\tau))d\tau \\ \tfrac{1-\xi}{ABC(\xi)}\mathfrak{F}(t) + \frac{\xi}{ABC(\xi)\Gamma(\xi)}\int_0^t (t-\tau)^{\xi-1}\mathfrak{F}(\tau)d\tau, \end{aligned}\right\} \tag{31}$$

$$\left.\begin{aligned}&\left|\left|\aleph(t)-\left(\aleph_0(t)+\tfrac{1-\xi}{ABC(\xi)}\eth(t,\aleph(t))+\tfrac{\xi}{ABC(\xi)\Gamma(\xi)}\int_0^t(t-\tau)^{\xi-1}\eth(\tau,\aleph(\tau))d\tau\right)\right|\right|\\&=\left|\left|\tfrac{1-\xi}{ABC(\xi)}\mathfrak{F}(t)+\tfrac{\xi}{ABC(\xi)\Gamma(\xi)}\int_0^t(t-\tau)^{\xi-1}\mathfrak{F}(\tau)d\tau\right|\right|.\end{aligned}\right\}\tag{32}$$

After simplification, we have

$$\left.\begin{aligned}&\left|\left|\aleph(t)-\left(\aleph_0(t)+\tfrac{1-\xi}{ABC(\xi)}\eth(t,\aleph(t))+\tfrac{\xi}{ABC(\xi)\Gamma(\xi)}\int_0^t(t-\tau)^{\xi-1}\eth(\tau,\aleph(\tau))d\tau\right)\right|\right|\\&\leq\epsilon\mathfrak{U},\end{aligned}\right\}$$

where $\mathfrak{U}=\frac{(\Gamma(\xi)(1-\xi)+T^{\xi})}{ABC(\xi)\Gamma(\xi)}$. □

Theorem 5. Assuming condition (II) and utilizing (30), the solution of the system (11) is demonstrated to exhibit UH stability. As a result, the analytical solution of the proposed system is confirmed to be UH stable when $\mathfrak{T}<1$, where $\mathfrak{T}=\mathfrak{W}_3\mathfrak{U}$.

Proof. Assuming that condition (II) is satisfied and that a unique solution exists, denoted as $\overline{\aleph}\in\mathbb{P}$ and considering any other solution $\aleph\in\mathbb{P}$ to (16), it follows that

$$\begin{aligned}&\left|\left|\aleph(t)-\overline{\aleph}(t)\right|\right|\\&=\left|\left|\aleph(t)-\left(\aleph_0(t)+\frac{1-\xi}{ABC(\xi)}\eth(t,\overline{\aleph}(t))\right.\right.\right.\\&\quad\left.\left.\left.+\frac{\xi}{ABC(\xi)\Gamma(\xi)}\int_0^t(t-\tau)^{\xi-1}\eth(\tau,\overline{\aleph}(\tau))d\tau\right)\right|\right|.\end{aligned}\tag{33}$$

$$\left.\begin{aligned}&\left|\left|\aleph(t)-\overline{\aleph}(t)\right|\right|\\&\leq\left|\left|\aleph(t)-\left(\aleph_0(t)+\tfrac{1-\xi}{ABC(\xi)}\eth(t,\aleph(t))\right.\right.\right.\\&\quad\left.\left.+\tfrac{\xi}{ABC(\xi)\Gamma(\xi)}\int_0^t(t-\tau)^{\xi-1}\eth(\tau,\aleph(\tau))d\tau\right)\right|\right|\\&\quad+\left|\left|\left(\aleph_0(t)+\tfrac{1-\xi}{ABC(\xi)}\eth(t,\aleph(t))+\tfrac{\xi}{ABC(\xi)\Gamma(\xi)}\int_0^t(t-\tau)^{\xi-1}\eth(\tau,\aleph(\tau))d\tau\right)\right.\right.\\&\quad\left.\left.-\left(\aleph_0(t)+\tfrac{1-\xi}{ABC(\xi)}\eth(t,\overline{\aleph}(t))+\tfrac{\xi}{ABC(\xi)\Gamma(\xi)}\int_0^t(t-\tau)^{\xi-1}\eth(\tau,\overline{\aleph}(\tau))d\tau\right)\right|\right|.\end{aligned}\right\}\tag{34}$$

$$\left|\left|\aleph(t)-\overline{\aleph}(t)\right|\right|\leq\frac{\mathfrak{U}\epsilon}{1-\mathfrak{T}},\tag{35}$$

where $\mathfrak{T} = \mathfrak{W}_3\mathfrak{U}$.

From (34), it follows that the solution of (11) is stable in the UH sense and, more generally, in the generalized UH sense, when $\mathfrak{Q}(\epsilon) = \mathfrak{U}\epsilon$ and $\mathfrak{Q}(0) = 0$. □

6 Disease-free equilibrium point and the basic reproduction number

The calculation of the equilibrium points (EPs) for the model described in (11) is outlined below. The equilibrium states represent points where the population sizes remain constant. First, we determine the EPs, followed by an analysis of the model's stability. The equilibrium states are obtained by solving the following equations:

$$\begin{cases} {}^{ABC}_{0}D_t^{\xi}\mathcal{X}(t) = 0, \\ {}^{ABC}_{0}D_t^{\xi}\mathcal{Y}(t) = 0, \\ {}^{ABC}_{0}D_t^{\xi}\mathcal{Z}(t) = 0. \end{cases} \tag{36}$$

The disease-free equilibrium (DFE), denoted $\mathscr{P}^0 = \left(\mathcal{X}^0, \mathcal{Y}^0, \mathcal{Z}^0\right)$, refers to the state in which the illness is not present in the system. The DFE is given by the following expression:

$$\begin{cases} \mathcal{X}^0 &= \frac{\Lambda(1-p)}{\mu_1}, \\ \mathcal{Y}^0 &= \frac{\Lambda p}{\mu_2}, \\ \mathcal{Z}^0 &= 0. \end{cases} \tag{37}$$

The basic reproduction number, commonly referred to as $\mathfrak{R}_0$, is a key epidemiological parameter that indicates the average number of secondary infections caused by a single infected individual in a fully susceptible population. It is essential for understanding the potential for Pneumocystis Pneumonia to spread. To calculate $\mathfrak{R}_0$, the next-generation matrix method is used. As a result, $\mathfrak{R}_0$ can be expressed in the following form:

$$\mathfrak{R}_0 = \rho(\mathscr{F}\mathscr{V}^{-1}). \tag{38}$$

In this context, ρ represents the spectral radius of the matrix $\mathscr{F}\mathscr{V}^{-1}$, where $\mathscr{F}$ is the matrix of new infection terms and $\mathscr{V}$ is the matrix of transfer terms out of the infected compartment:

$$\mathscr{F} = \begin{pmatrix} \beta_1 \mathcal{X}^0 + \beta_2(1-\psi)\mathcal{Y}^0 \\ \mathcal{X}^0 + \mathcal{Y}^0 \end{pmatrix}, \qquad \mathscr{V} = \mu_3. \tag{39}$$

Therefore, $\mathfrak{R}_0$ can be represented as $\mathfrak{R}_0 = \rho(\mathscr{F}\mathscr{V}^{-1}) = \begin{pmatrix} \frac{\beta_1 \mathcal{X}^0 + \beta_2(1-\psi)\mathcal{Y}^0}{\mu_3(\mathcal{X}^0+\mathcal{Y}^0)} \end{pmatrix} = \frac{\beta_2\mu_1 p(1-\psi)+\beta_1\mu_2(1-p)}{\mu_3(\mu_1 p+\mu_2(1-p))}$.

The analytical expression obtained for the basic reproduction number $\mathfrak{R}_0$ reveals that it is governed by several key parameters of the model. A value of $\mathfrak{R}_0 > 1$ indicates that there is a transmission of the disease between the immune-deficient population that can be sustained. This is because on average an infectious individual gives rise to more than one secondary case. However, the situation is reversed when $\mathfrak{R}_0 < 1$ and the disease will slowly disappear in the population. The magnitude of $\mathfrak{R}_0$ is affected by multiple factors, including the average duration of infectiousness (associated with $\frac{1}{\mu_3}$), the transmission coefficients β_1 and β_2, as well as control-related parameters such as prophylaxis uptake p and its effectiveness ψ. Owing to its central role in epidemiological analysis, $\mathfrak{R}_0$ serves as a fundamental indicator for evaluating intervention strategies and informing public health decision-making, particularly for protecting high-risk populations vulnerable to Pneumocystis pneumonia.

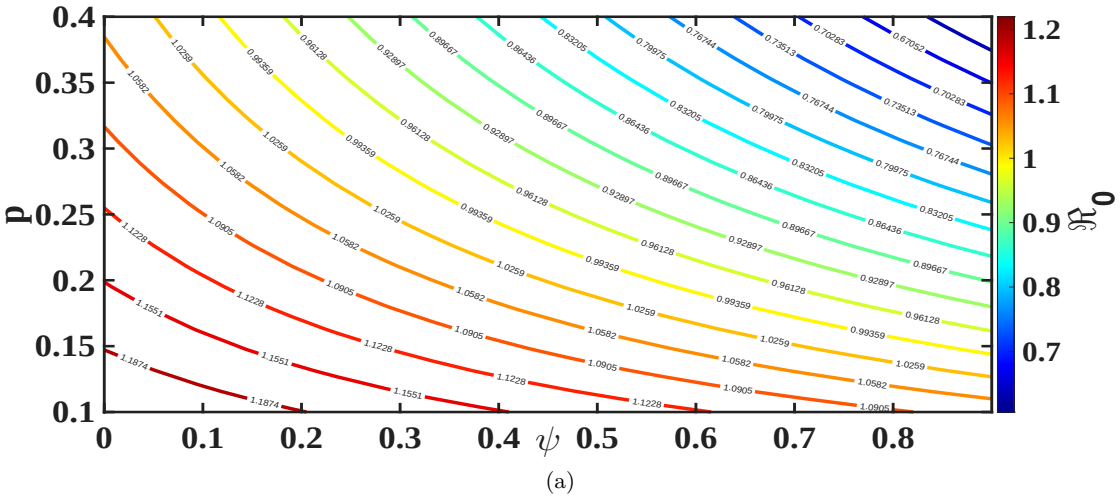
We investigate the effect of two-parameter variations on the value of the basic reproduction number. Figures 2(a) and 2(b) show that prophylaxis efficacy and prophylaxis coverage lead to $\mathfrak{R}_0 < 1$, indicating the disease-free state in the region

6.1 Existence of endemic equilibrium point

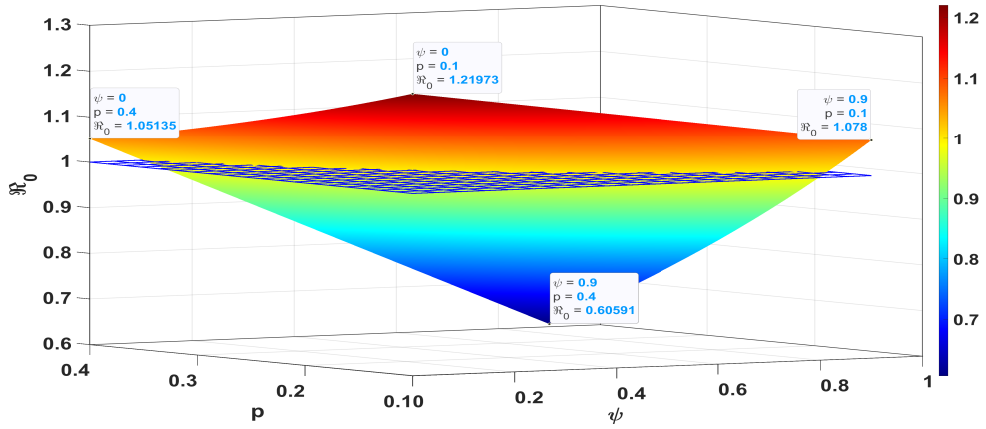
The endemic equilibrium $\mathscr{P}^* = (\mathcal{X}^*, \mathcal{Y}^*, \mathcal{Z}^*)$ corresponds to a steady state in which Pneumocystis Pneumonia remains present within the population, characterized by a strictly positive level of infection. This scenario is of high concern, as it reflects sustained transmission. At steady state, we can consider $\mathcal{N}^* = \mathcal{X}^* + \mathcal{Y}^* + \mathcal{Z}^* = \mathcal{X}^0 + \mathcal{Y}^0 + \mathcal{Z}^0 = \frac{\Lambda(1-p)\mu_2+\Lambda p}{\mu_1\mu_2}$ which is a constant total population. The components of the EEP can be expressed as follows:

$$\begin{cases} \mathcal{X}^* = \frac{\Lambda(1-p)\mathcal{N}^*}{\beta_1\mathcal{Z}^*+\mu_1\mathcal{N}^*}, \\ \mathcal{Y}^* = \frac{\Lambda p\mathcal{N}^*}{\beta_2(1-\psi)\mathcal{Z}^*+\mu_2\mathcal{N}^*}. \end{cases} \tag{40}$$

To determine the positive solution $\mathcal{Z}^*$, the expressions for $\mathcal{X}^*$ and $\mathcal{Y}^*$ given in (40), together with $\mathcal{Z}^* = \mathcal{N}^* - \mathcal{X}^* + \mathcal{Y}^*$, are substituted into the third equation of system (11). This substitution leads to a quadratic equation in $\mathcal{Z}^*$,



(a)



(b)

Figure 2: The impact of parameters ψ and p on $\mathfrak{R}_0$

$$\mathcal{Q}(\mathcal{Z}^*) = \mathscr{A}(\mathcal{Z}^*)^2 + \mathscr{B}(\mathcal{Z}^*) + \mathscr{C}, \tag{41}$$

where

$$\begin{cases} \mathscr{A} = \beta_1\beta_2\mu_3(1-\psi), \\ \mathscr{B} = \beta_2(1-\psi)(\mu_3\mu_1\mathcal{N}^* - \beta_1\Lambda) + \beta_1\mu_2\mu_3\mathcal{N}^*, \\ \mathscr{C} = \mu_1\mu_2\mu_3(\mathcal{N}^*)^2 + \mu_1\Lambda\beta_2 p\mathcal{N}^*(\psi-1) + \beta_1\mu_2\Lambda(p-1)\mathcal{N}^* \\ \quad = \mu_1\mu_2\mu_3\mathcal{N}^*(\mathcal{X}^0+\mathcal{Y}^0)\left(\frac{\mathcal{N}^*}{(\mathcal{X}^0+\mathcal{Y}^0)} - \mathfrak{R}_0\right). \end{cases} \tag{42}$$

An endemic equilibrium exists if there is at least one positive real root for $\mathscr{Z}^*$. The existence of such a root depends on the signs of the coefficients $\mathscr{A}$, $\mathscr{B}$, and $\mathscr{C}$, it is easily seen that $\mathscr{A}$ always positive and $\mathscr{C}$ is negative whenever $\mathfrak{R}_0 > 1$. By the Descartes rule of signs, the polynomial (41) has a positive root, and an endemic equilibrium becomes feasible when $\mathfrak{R}_0 > 1$.

6.2 Local stability of equilibria

In the following part, we devote our attention to examining the stability of the EPs. Stability analysis determines whether the population will return to an equilibrium state after a small disturbance, such as the introduction of a few infected individuals. To begin, we present the theorem below which is fundamental for fractional-order systems.

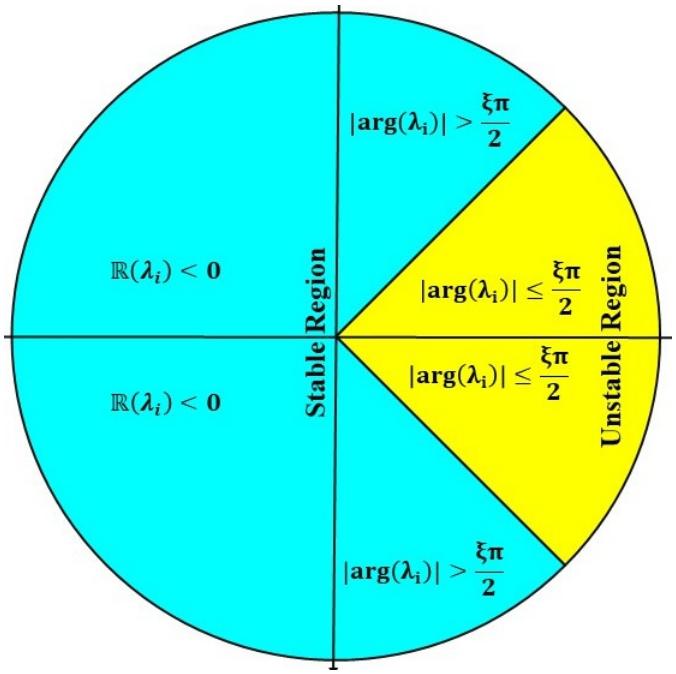


Figure 3: Stable region for fractional-order model.

Figure 3 displays the regions of stability (indicated in blue) and instability (indicated in yellow) for a fractional-order system defined by the order ξ. Figure 3 shows that the stability region for the fractional-order system is considerably larger than that of an integer-order system. Specifically, the integer-order system is stable only if all eigenvalues have negative real parts, which covers parts of the second and third quadrants. In contrast, the fractional-order system

exhibits a more extensive stability domain that extends beyond this axis. Therefore, while the stability of the integer-order model guarantees stability for the corresponding fractional-order model, the reverse is not true; a stable fractional-order model does not ensure the stability of the integer-order version.

Theorem 6. We begin by examining the following system of fractional order:

$$ {}^{ABC}_{0}D_t^{\xi}\mathcal{F}(t) = \mathscr{M}(t, \mathcal{F}(t)). \tag{43}$$

Suppose that $\mathcal{F}^*$ is an EP of the system given in (43), with $\mathcal{J}(\mathcal{F}^*)$ representing the Jacobian matrix evaluated at $\mathcal{F}^*$, $\lambda_i (i = 1, 2, \ldots, n)$ denoting its eigenvalues. Under these conditions, the system satisfies the following criteria:

(a) The EP $\mathcal{F}^*$ is locally asymptotically stable if each eigenvalue λ_i satisfies the condition $|arg(\lambda_i)| > \frac{\xi\pi}{2}$.

(b) The EP $\mathcal{F}^*$ is deemed unstable if there exists at least one eigenvalue λ_k, for some $k \in \{1, 2, \ldots, n\}$.

Next, we proceed to analyze the local stability of each EP individually. The Jacobian matrix corresponding to the DFE point is expressed in the following form:

$$\mathcal{J}(\mathscr{P}^0) = \begin{pmatrix} -\mu_1 & 0 & -\frac{\beta_1(\beta_2(\psi-1)+\mu_3\mathfrak{R}_0)}{\beta_2(\psi-1)+\beta_1} \\ 0 & -\mu_2 & \frac{\beta_2(\psi-1)(\beta_1-\mu_3\mathfrak{R}_0)}{\beta_2(\psi-1)+\beta_1} \\ 0 & 0 & \mu_3(\mathfrak{R}_0-1) \end{pmatrix}. \tag{44}$$

The eigenvalues of $\mathcal{J}(\mathscr{P}^0)$ are $\lambda_1 = -\mu_1$, $\lambda_2 = -\mu_2$ and $\lambda_3 = \mu_3(\mathfrak{R}_0 - 1)$. It is easily seen that eigenvalues are negative if and only if $\mathfrak{R}_0 < 1$, establishing that the DFE $\mathscr{P}^0$ is locally asymptotically stable.

The Jacobian matrix evaluated at the endemic EPs $\mathscr{P}^*$ is given by

$$\mathcal{J}(\mathscr{P}^*) = \begin{pmatrix} -\frac{\beta_1\mathcal{Z}^*(\mathcal{Z}^*+\mathcal{Y}^*)}{(\mathcal{X}^*+\mathcal{Y}^*+\mathcal{Z}^*)^2} - \mu_1 & \frac{\beta_1\mathcal{X}^*\mathcal{Z}^*}{(\mathcal{X}^*+\mathcal{Y}^*+\mathcal{Z}^*)^2} & -\frac{\beta_1\mathcal{X}^*(\mathcal{X}^*+\mathcal{Y}^*)}{(\mathcal{X}^*+\mathcal{Y}^*+\mathcal{Z}^*)^2} \\ -\frac{\beta_2(\psi-1)\mathcal{Z}^*\mathcal{Y}^*}{(\mathcal{Z}^*+\mathcal{Z}^*+\mathcal{Z}^*)^2} & \frac{\beta_2(\psi-1)\mathcal{Z}^*(\mathcal{X}^*+\mathcal{Z}^*)}{(\mathcal{X}^*+\mathcal{Y}^*+\mathcal{Z}^*)^2} - \mu_2 & \frac{\beta_2(\psi-1)\mathcal{Y}^*(\mathcal{X}^*+\mathcal{Y}^*)}{(\mathcal{X}^*+\mathcal{Y}^*+\mathcal{Z}^*)^2} \\ \frac{\mathcal{Z}^*(\beta_1(\mathcal{Z}^*+\mathcal{Y}^*)+\beta_2(\psi-1)\mathcal{Y}^*)}{(\mathcal{X}^*+\mathcal{Y}^*+\mathcal{Z}^*)^2} & -\frac{\mathcal{Z}^*(\beta_2(\psi-1)(\mathcal{X}^*+\mathcal{Z}^*)+\beta_1\mathcal{Z}^*)}{(\mathcal{X}^*+\mathcal{Y}^*+\mathcal{Z}^*)^2} & \frac{-\mu_3(\mathcal{X}^*+\mathcal{Y}^*+\mathcal{Z}^*)^2-\beta_2(\psi-1)\mathcal{Y}^*(\mathcal{X}^*+\mathcal{V}^*)+\beta_1}{(\mathcal{X}^*+\mathcal{Y}^*+\mathcal{Z}^*)^2} \end{pmatrix}, \tag{45}$$

$$\mathcal{J}(\mathscr{P}^*) = \begin{pmatrix} j_{11} & j_{12} & j_{13} \\ j_{21} & j_{22} & j_{23} \\ j_{31} & j_{32} & j_{33} \end{pmatrix}.$$

The characteristic equation of $\mathcal{J}(\mathscr{P}^*)$ is given by

$$\lambda^3 + p_1\lambda^2 + p_2\lambda + p_3 = 0, \tag{46}$$

where

$$\begin{aligned} p_1 &= j_{11} + j_{22} + j_{33}, \\ p_2 &= j_{22}j_{33} - j_{32}j_{23} + j_{11}j_{33} - j_{31}j_{13} + j_{11}j_{22} - j_{22}j_{12}, \\ p_3 &= j_{11}(j_{22}j_{33} - j_{32}j_{23}) - j_{12}(j_{22}j_{33} - j_{31}j_{23}) + j_{13}(j_{22}j_{32} - j_{31}j_{22}). \end{aligned}$$

If $p_1 > 0, p_2 > 0$, and $p_3 > 0$ with $p_1p_2 - p_3 > 0$, then all eigenvalues of $\mathcal{J}(\mathscr{P}^*)$ would lie in the left half-plane; hence the endemic equilibrium is locally asymptotically stable.

7 Sensitivity index

This section explores how variations in these parameters affect the transmission dynamics of the disease through sensitivity analysis. This process identifies which parameters have the most significant impact on $\mathfrak{R}_0$, thereby highlighting the most effective targets for control interventions. The parameter space is defined as

$$\mathbb{W} = \{(p, \beta_1, \beta_2, \psi, \mu_1, \mu_2, \mu_3) \in \mathbb{R}_+^7\}, \tag{47}$$

where $\mathfrak{R}_0 \in C^1(\mathbb{W})$. To evaluate how variations in a parameter $\eta \in \mathbb{W}$ influence $\mathfrak{R}_0$, we introduce the normalized forward sensitivity index, defined as

$$\mathcal{A}_\eta^{\mathfrak{R}_0} = \frac{\partial \mathfrak{R}_0}{\partial \eta} \times \frac{\eta}{\mathfrak{R}_0}. \tag{48}$$

This index quantifies the percentage change in $\mathfrak{R}_0$ resulting from a 1% change in the parameter η. A positive index means $\mathfrak{R}_0$ increases as the parameter increases, while a negative index

means that $\mathfrak{R}_0$ decreases. This analysis is crucial for prioritizing public health strategies for Pneumocystis Pneumonia control. This leads us to the following results:

$$\begin{aligned}
\mathcal{A}_p^{\mathfrak{R}_0} &= \frac{\partial \mathfrak{R}_0}{\partial p} \times \frac{p}{\mathfrak{R}_0} \\
&= \frac{p\mu_3(p\mu_1+\mu_2(1-p))\left(\frac{\mu_1\beta_2(1-psi)-\beta_1\mu_2}{\mu_3(\mu_1 p+\mu_2(1-p))} - \frac{(\beta_2\mu_1 p(1-psi)+\beta_1\mu_2(1-p))(\mu_1-\mu_2)}{\mu_3(\mu_1 p+\mu_2(1-p))^2}\right)}{p\mu_1\beta_2(1-\psi)+\beta_1\mu_2(1-p)}, \\
\mathcal{A}_{\beta_1}^{\mathfrak{R}_0} &= \frac{\partial \mathfrak{R}_0}{\partial \beta_1} \times \frac{\beta_1}{\mathfrak{R}_0} \\
&= \frac{\beta_1\mu_2(1-p)}{\beta_2\mu_1 p(1-\psi)+\beta_1\mu_2(1-p)}, \\
\mathcal{A}_{\beta_2}^{\mathfrak{R}_0} &= \frac{\partial \mathfrak{R}_0}{\partial \beta_2} \times \frac{\beta_2}{\mathfrak{R}_0} \\
&= \frac{\beta_2\mu_1 p(1-\psi)}{\beta_2\mu_1 p(1-\psi)+\beta_1\mu_2(1-p)}, \\
\mathcal{A}_{\psi}^{\mathfrak{R}_0} &= \frac{\partial \mathfrak{R}_0}{\partial \psi} \times \frac{\psi}{\mathfrak{R}_0} \\
&= -\frac{\beta_2\mu_1 p\psi}{\beta_2\mu_1 p(1-\psi)+\beta_1\mu_2(1-p)}, \\
\mathcal{A}_{\mu_1}^{\mathfrak{R}_0} &= \frac{\partial \mathfrak{R}_0}{\partial \mu_1} \times \frac{\mu_1}{\mathfrak{R}_0} \\
&= \frac{\mu_1\mu_3(\mu_1 p+\mu_2(1-p))\left(\frac{p\beta_2(1-\psi)}{\mu_3(\mu_1 p+\mu_2(1-p))} - \frac{p(\beta_2\mu_1 p(1-\psi)+\beta_1\mu_2(1-p))}{\mu_3(\mu_1 p+\mu_2(1-p))^2}\right)}{\beta_2\mu_1 p(1-\psi)+\beta_1\mu_2(1-p)}, \\
\mathcal{A}_{\mu_2}^{\mathfrak{R}_0} &= \frac{\partial \mathfrak{R}_0}{\partial \mu_2} \times \frac{\mu_2}{\mathfrak{R}_0} \\
&= \frac{\mu_2\mu_3(\mu_1 p+\mu_2(1-p))\left(\frac{\beta_1(1-p)}{\mu_3(\mu_1 p+\mu_2(1-p))} - \frac{p(\beta_2\mu_1 p(1-\psi)+\beta_1\mu_2(1-p))}{\mu_3(\mu_1 p+\mu_2(1-p))^2}\right)}{\beta_2\mu_1 p(1-\psi)+\beta_1\mu_2(1-p)}, \\
\mathcal{A}_{\mu_3}^{\mathfrak{R}_0} &= \frac{\partial \mathfrak{R}_0}{\partial \mu_3} \times \frac{\mu_3}{\mathfrak{R}_0} = -1.
\end{aligned}$$

The sensitivity of the basic reproduction number with respect to the model parameters is illustrated in Figure 4, where the normalized sensitivity indices identify the parameters that have the greatest influence on disease transmission.

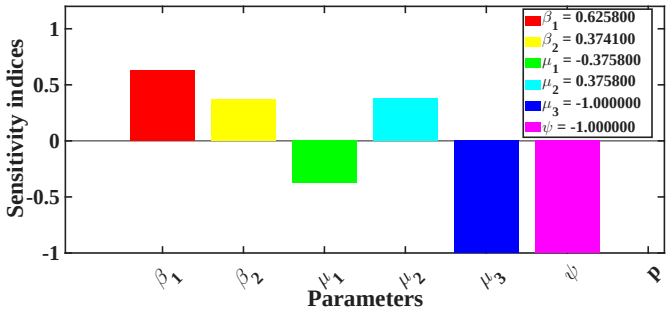


Figure 4: Sensitivity indices of $\Re_0$

8 Numerical scheme

The numerical simulations of the proposed fractional-order pneumonia model are performed using the numerical method developed by Solís-Pérez et al. [40]. This method is based on two-step Lagrange polynomial interpolation and is applied to the ABC fractional system (11) to investigate its dynamical behavior. An error and convergence analysis is also carried out to establish the convergence behavior of the implemented numerical scheme. This approach is as follows:

$$ {}^{ABC}_{0}D_t^{\xi}\aleph(t) = \eth(t, \aleph(t)). \tag{49}$$

An equivalent integral representation of the fractional-order differential equation (49) is derived via the fundamental theorem of fractional calculus.

$$\aleph(t) - \aleph(0) = \frac{1-\xi}{ABC(\xi)}\eth(t, \aleph(t)) + \frac{\xi}{ABC(\xi)\Gamma(\xi)}\int_0^t (t-\tau)^{\xi-1}\eth(\tau, \aleph(\tau))d\tau, \tag{50}$$

where $ABC(\xi) = 1 - \xi + \frac{\xi}{\Gamma(\xi)}$ is the normalization function. Equation (50) can be rewritten as

$$\begin{aligned}\aleph(t_{j+1}) - \aleph(0) =& \frac{\Gamma(\xi)(1-\xi)}{\Gamma(\xi)(1-\xi)+\xi}\eth(t_j, \aleph(t_j)) \\ &+ \frac{\xi}{\Gamma(\xi)+\xi(1-\Gamma(\xi))}\sum_{n=0}^{j}\int_{t_n}^{t_{n+1}} \eth(\tau, \aleph(\tau)(t_{n+1}-\tau)^{\tau-1}d\tau.\end{aligned} \tag{51}$$

The function $\eth(\tau, \aleph(\tau))$ has an approximate value obtained by the two-step Lagrange polynomial interpolation. Throughout the range $[t_n, t_{n+1}]$.

$$\mathfrak{G}_m(\tau) \simeq \frac{\eth(t_n, \aleph_n)}{h}(\tau - t_{n-1}) - \frac{\eth(t_{n-1}, \aleph_{n-1})}{h}(\tau - t_n). \tag{52}$$

Now, taking into account (52) and (51), we arrive at

$$\left.\begin{aligned} \aleph_{j+1}(t) = \; & \aleph_0 + \tfrac{(1-\xi)\Gamma(\xi)}{\xi+(1-\xi)\Gamma(\xi)}\eth(t_j, \aleph(t_j)) + \tfrac{\xi}{\Gamma(\xi)+\xi(1-\Gamma(\xi))} \\ & + \sum_{n=0}^{j}\Bigg(\frac{\eth(t_n, \aleph_n)}{h}\int_{t_n}^{t_{n+1}}(\tau - t_{n-1})(t_{j+1} - \tau)^{\xi-1}d\tau \\ & - \tfrac{\eth(t_{n-1}, \aleph_{n-1})}{h}\int_{t_n}^{t_{n+1}}(\tau - t_n)(t_{j+1} - \tau)^{\xi-1}d\tau\Bigg). \end{aligned}\right\} \tag{53}$$

For simplicity, we let

$$\left.\begin{aligned} \mathfrak{B}_{\xi,n,1} &= \int_{t_n}^{t_{n+1}}(\tau - t_{n-1})(t_{j+1} - \tau)^{\xi-1}d\tau, \\ \mathfrak{B}_{\xi,n,2} &= \int_{t_n}^{t_{n+1}}(\tau - t_n)(t_{j+1} - \tau)^{\xi-1}d\tau, \end{aligned}\right\} \tag{54}$$

which implies that

$$\left.\begin{aligned} \mathfrak{B}_{\xi,n,1} &= h^{\xi+1}\tfrac{(j+1-n)^{\xi}(j-n+2+\xi)-(j-n)^{\xi}(j-n+2-2\xi)}{\xi(\xi+1)}, \\ \mathfrak{B}_{\xi,n,2} &= h^{\xi+1}\tfrac{(j+1-n)^{j+1}-(j-n)^{\xi}(j-n+1+\xi)}{\xi(\xi+1)}. \end{aligned}\right\} \tag{55}$$

Substituting the integration (55) into (53), provides a viable approximation

$$\left.\begin{aligned} \aleph_{j+1}(t) = \; & \aleph_0 + \tfrac{(1-\xi)\Gamma(\xi)}{\xi+(1-\xi)\Gamma(\xi)}\eth(t_j, \aleph(t_j)) + \tfrac{1}{\xi+(1+\xi)(1-\xi)\Gamma(\xi)} \\ & \times\sum_{n=0}^{j}\Bigg(h^{\xi}\eth(t_n, \aleph_n)\big((j+1-n)^{\xi}(j-n+2+\xi) \\ & -(j-n)^{\xi}(j-n+2+2\xi)\big) - h^{\xi}\eth(t_{n-1}, \aleph_{n-1})\big((j+1-n)^{\xi+1} \\ & -(j-n)^{\xi}\times(j-n+1+\xi)\big)\Bigg). \end{aligned}\right\} \tag{56}$$

Finally, the numerical solution of model (11) is found as follows:

$$\left.\begin{aligned}\mathcal{X}_{j+1}(t) = \;& \mathcal{X}_0 + \tfrac{(1-\xi)\Gamma(\xi)}{\xi+(1-\xi)\Gamma(\xi)}\aleph_1(t_j,\mathcal{X}(t_j),\mathcal{Y}(t_j),\mathcal{Z}(t_j)) + \tfrac{1}{\xi+(1+\xi)(1-\xi)\Gamma(\xi)}\\ &\times\sum_{n=0}^{j}\Big(h^{\xi}\aleph_1(t_n,\mathcal{X}_n,\mathcal{Y}_n,\mathcal{Z}_n)\big((j+1-n)^{\xi}(j-n+2+\xi)\\ &-(j-n)^{\xi}(j-n+2+2\xi)\big)\\ &-h^{\xi}\aleph_1(t_{n-1},\mathcal{X}_{n-1},\mathcal{Y}_{n-1},\mathcal{Z}_{n-1})\big((j+1-n)^{\xi+1}\\ &-(j-n)^{\xi}\times(j-n+1+\xi)\big)\Big),\end{aligned}\right\} \tag{57}$$

$$\left.\begin{aligned}\mathcal{Y}_{j+1}(t) = \;& \mathcal{Y}_0 + \tfrac{(1-\xi)\Gamma(\xi)}{\xi+(1-\xi)\Gamma(\xi)}\aleph_2(t_j,\mathcal{X}(t_j),\mathcal{Y}(t_j),\mathcal{Z}(t_j)) + \tfrac{1}{\xi+(1+\xi)(1-\xi)\Gamma(\xi)}\\ &\times\sum_{n=0}^{j}\Big(h^{\xi}\aleph_2(t_n,\mathcal{X}_n,\mathcal{Y}_n,\mathcal{Z}_n)\big((j+1-n)^{\xi}(j-n+2+\xi)\\ &-(j-n)^{\xi}(j-n+2+2\xi)\big)\\ &-h^{\xi}\aleph_2(t_{n-1},\mathcal{X}_{n-1},\mathcal{Y}_{n-1},\mathcal{Z}_{n-1})\big((j+1-n)^{\xi+1}\\ &-(j-n)^{\xi}\times(j-n+1+\xi)\big)\Big),\end{aligned}\right\} \tag{58}$$

$$\left.\begin{aligned}\mathcal{Z}_{j+1}(t) = \;& \mathcal{Z}_0 + \tfrac{(1-\xi)\Gamma(\xi)}{\xi+(1-\xi)\Gamma(\xi)}\aleph_3(t_j,\mathcal{X}(t_j),\mathcal{Y}(t_j),\mathcal{Z}(t_j)) + \tfrac{1}{\xi+(1+\xi)(1-\xi)\Gamma(\xi)}\\ &\times\sum_{n=0}^{j}\Big(h^{\xi}\aleph_3(t_n,\mathcal{X}_n,\mathcal{Y}_n,\mathcal{Z}_n)\big((j+1-n)^{\xi}(j-n+2+\xi)\\ &-(j-n)^{\xi}(j-n+2+2\xi)\big)\\ &-h^{\xi}\aleph_3(t_{n-1},\mathcal{X}_{n-1},\mathcal{Y}_{n-1},\mathcal{Z}_{n-1})\big((j+1-n)^{\xi+1}\\ &-(j-n)^{\xi}\times(j-n+1+\xi)\big)\Big).\end{aligned}\right\} \tag{59}$$

8.1 Error analysis

We now present the error analysis to validate the proposed scheme. Assuming that the function $\eth(t,\aleph(t))$ is twice differentiable, our given problem is as follows:

$$\begin{cases} {}^{ABC}_{0}D_t^{\xi}\aleph(t) = \eth(t,\aleph(t)), & t\in K=[0,T],\\ \aleph(0)=\aleph_0, & 0<\xi\le 1.\end{cases}$$

The operator involved possesses a nonlocal and nonsingular kernel. Assuming that the function has a bounded second derivative, (16) can be reformulated as follows:

$$\begin{aligned}
\aleph(t_{k+1}) - \aleph(0) =& \frac{1-\xi}{ABC(\xi)}\eth(t_k, \aleph(t_k)) + \frac{\xi}{ABC(\xi)\Gamma(\xi)} \\
&\times \int_0^{t_{k+1}} \eth(\mathfrak{y}, \aleph(\mathfrak{y}))(t_{k+1}-\mathfrak{y})^{\xi-1} d\mathfrak{y} \\
=& \frac{1-\xi}{ABC(\xi)}\eth(t_k, \aleph(t_k)) + \frac{\xi}{ABC(\xi)\Gamma(\xi)} \\
&\times \sum_{n=0}^{k} \int_{t_n}^{t_{n+1}} \eth(\mathfrak{y}, \aleph(\mathfrak{y}))(t_{k+1}-\mathfrak{y})^{\xi-1} d\mathfrak{y} \\
=& \frac{1-\xi}{ABC(\xi)}\eth(t_k, \aleph(t_k)) + \frac{\xi}{ABC(\xi)\Gamma(\xi)} \\
&\times \sum_{n=0}^{k} \int_{t_n}^{t_{n+1}} \left\{ P_n(\mathfrak{y}) + \frac{(\mathfrak{y}-t_n)(\mathfrak{y}-t_{n-1})}{2!} \frac{\partial^2}{\partial \mathfrak{y}^2}[\eth(\mathfrak{y}, \aleph(\mathfrak{y}))]_{\mathfrak{y}=\gamma_n} \right\} \\
&\times (t_{k+1}-\mathfrak{y})^{\xi-1} d\mathfrak{y} \frac{1-\xi}{ABC(\xi)}\eth(t_k, \aleph(t_k)) + E_k^{\xi} + \frac{\xi}{ABC(\xi)\Gamma(\xi)} \\
&\sum_{n=0}^{k} \left\{ \frac{\eth(t_n, \aleph^n)}{h} \int_{t_n}^{t_{n+1}} (\mathfrak{y}-t_{n-1})(t_{k+1}-\mathfrak{y})^{\xi-1} d\mathfrak{y} \right. \\
&\left. - \frac{\eth(t_{n-1}, \aleph^{n-1})}{h} \int_{t_n}^{t_{n+1}} (\mathfrak{y}-t_n)(t_{k+1}-\mathfrak{y})^{\xi-1} d\mathfrak{y} \right\},
\end{aligned} \tag{60}$$

where

$$\begin{aligned}
E_k^{\xi} =& \frac{\xi}{ABC(\xi)\Gamma(\xi)} \sum_{n=0}^{k} \int_{t_n}^{t_{n+1}} \frac{(\mathfrak{y}-t_n)(\mathfrak{y}-t_{n-1})}{2!} \\
&\times \frac{\partial^2}{\partial \mathfrak{y}^2}[\eth(\mathfrak{y}, \aleph(\mathfrak{y}))]_{\mathfrak{y}=\gamma_n} (t_{k+1}-\mathfrak{y})^{\xi-1} d\mathfrak{y}.
\end{aligned} \tag{61}$$

One can find a point $\gamma_n \in [t_n, t_{n+1}]$ such that

$$\begin{aligned}
E_k^{\xi} =& \frac{\xi}{ABC(\xi)\Gamma(\xi)} \sum_{n=0}^{k} \frac{\partial^2}{\partial \mathfrak{y}^2}[\eth(\mathfrak{y}, \aleph(\mathfrak{y}))]_{\mathfrak{y}=\gamma_n} \frac{(\gamma_n - t_n)}{2} \\
&\times \int_{t_n}^{t_{n+1}} (\mathfrak{y}-t_{n-1})(t_{k+1}-\mathfrak{y})^{\xi-1} d\mathfrak{y} \\
=& \frac{\xi}{ABC(\xi)\Gamma(\xi)} \sum_{n=0}^{k} \frac{\partial^2}{\partial \mathfrak{y}^2}[\eth(\mathfrak{y}, \aleph(\mathfrak{y}))]_{\mathfrak{y}=\gamma_n} \frac{(\gamma_n - t_n)}{2} h^{\xi+1} \\
&\times \frac{1}{\xi(\xi+1)} \Big[(k-n+1)^{\xi}(k-n+2+\xi) - (k-n)^{\xi}(k-n+2+2\xi) \Big].
\end{aligned} \tag{62}$$

Taking the norm on both sides of (62), we obtain the following inequality:

$$||E_k^{\xi}|| \leq \frac{\xi h^{\xi+2}}{2ABC(\xi)\Gamma(\xi+2)} \underset{\mathfrak{y}\in[0,t_{k+1}]}{Sup}\left|\frac{\partial^2}{\partial \mathfrak{y}^2}\eth(\mathfrak{y},\aleph(\mathfrak{y}))\right| \times \sum_{n=0}^{k}\left|\left[(k-n+1)^{\xi}(k-n+2+\xi)-(k-n)^{\xi}(k-n+2+2\xi)\right]\right|. \tag{63}$$

Thus, we can calculate the error as follows:

$$||E_k^{\xi}|| \leq \frac{\xi h^{\xi+2}}{2ABC(\xi)\Gamma(\xi+2)} \underset{\mathfrak{y}\in[0,t_{k+1}]}{Sup}\left|\frac{\partial^2}{\partial \mathfrak{y}^2}\eth(\mathfrak{y},\aleph(\mathfrak{y}))\right| \times [(k+1)^{\xi}-\xi k^{\xi}]\frac{k(k+4+2\xi)}{2}. \tag{64}$$

From the derived error bound, it is evident that the numerical scheme exhibits a convergence rate of $\mathscr{O}(h^{\xi+2})$.

8.2 Computational experiments and results

Figures 5(a) and 5(b) show that increasing the fractional order leads to upward trends in the susceptible and prophylaxis-protected populations over time, respectively. However, Figure 5(c) reflects an opposite scenario in which a higher fractional order is associated with a decreasing trend in the infected population. This means that elevating the fractional order is beneficial for increasing the number of noninfected individuals. At the same time, it is a hindrance to the spread among the infected class. This behavior can be explained through the memory effect inherent in the Atangana–Baleanu fractional derivative. A smaller value of ξ corresponds to a stronger memory effect, implying that the influence of past infection states persists for a longer duration. Consequently, the force of infection from previously infected individuals continues to affect the susceptible and partially protected populations over an extended period. This prolonged transmission pressure results in sustained disease spread, leading to a higher number of infected individuals and relatively lower sizes of the susceptible and protected compartments. Conversely, as ξ approaches 1 (integer-order case), the memory effect weakens. The impact of past infections diminishes more rapidly, allowing the prophylactic intervention to act more effectively. The infection chain breaks quickly due to reduced lingering transmission pressure from previously infected cases. As a result, fewer individuals transition into the infected com-

partment, while more individuals remain in or move to the susceptible and prophylaxis-protected compartments. These findings indicate that the fractional-order model with higher ξ values better captures scenarios where prophylactic measures are more successful in controlling Pneumocystis pneumonia transmission within the immunocompromised population.

Figures 6(a), 6(b), and 6(c) show that the transmission rate from unprotected susceptibles to infected individuals (β_1) affects the population dynamics. When β_1 goes up, the sizes of the susceptible and prophylaxis-protected populations shrink while the infected population goes up drastically. The reason for this behavior is that β_1 directly increases the infection term in the model, thereby increasing the force of infection and speeding up the transition of individuals into the infected compartment. The findings here confirm that β_1 plays the primary role in disease spread within the fractional-order framework.

Figures 7(a), 7(b), and 7(c) show how the change in transmission rate from the prophylaxis-protected individuals to the infected ones (β_2) affects the population dynamics. When β_2 increases, the numbers of susceptible and prophylaxis-protected populations both decrease, whereas the infected population increases. The exposed risk increases, so the behavior of prophylactic protection being less effective is reflected here. The parameter β_2, which is the rate of spread of infection from partially protected individuals to susceptible individuals, determines the extent of the protective effect of prophylaxis. As β_2 increases, the impact of the protective measure decreases, while community transmission increases

Figures 8(a), 8(b), and 8(c) show that an increase in the efficacy of the prophylactic drug regimen (ψ) results in a growth of both the susceptible and prophylaxis-protected classes, along with a sharp decrease in the infected class. Formally, an increase in ψ reduces the effective infection term for partially protected individuals, thereby decreasing the total force of infection. From a biological standpoint, it means that enhanced prophylactic effectiveness increases community protection and limits the spread of the disease within the susceptible population.

Figures 9(a), 9(b), and 9(c) depict how the prophylaxis coverage rate for newly at-risk individuals (p) alters population dynamics. As more people are covered, the number of susceptibles decreases, and the number of individuals protected by prophylaxis increases. Moreover, the number of infected people is reduced dramatically. The reason for this change is that more and more newly recruited individuals receive prophylactic protection right away, resulting in fewer unprotected susceptibles and, consequently, a lower overall infection force. On the other hand, the higher the prophylaxis coverage, the better the disease is controlled in the community.

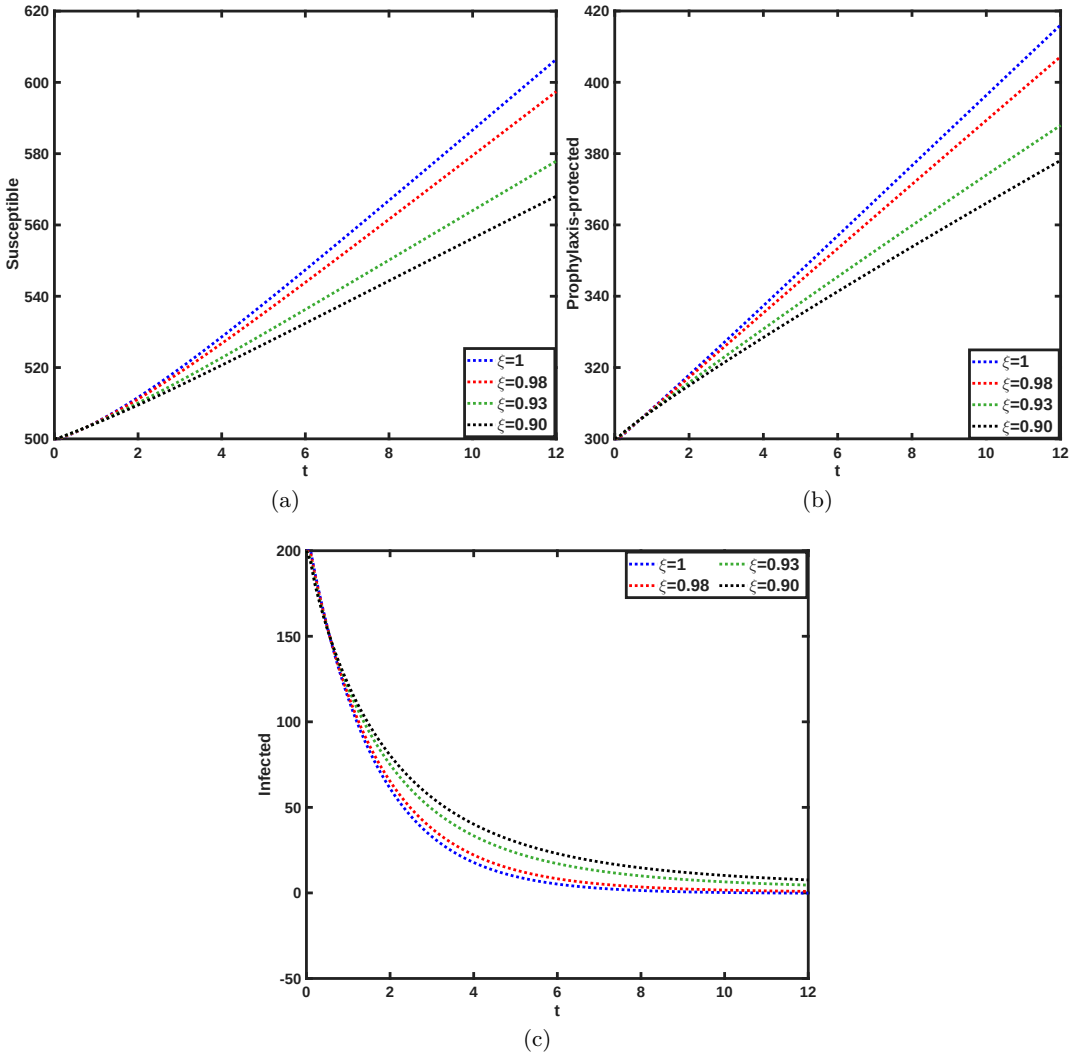


Figure 5: Effect of parameter ξ on population dynamics.

9 Conclusion

In the present study, a novel fractional-order mathematical model based on the Atangana–Baleanu–Caputo derivative was developed to study the transmission dynamics of pneumonia among the immunocompromised population. The model considers prophylactic protection as an imperfect preventive mechanism and, through memory effects in the fractional framework, provides a more realistic representation of disease progression than classical integer-order models. The theoretical analysis has established the positivity, boundedness, existence, and uniqueness

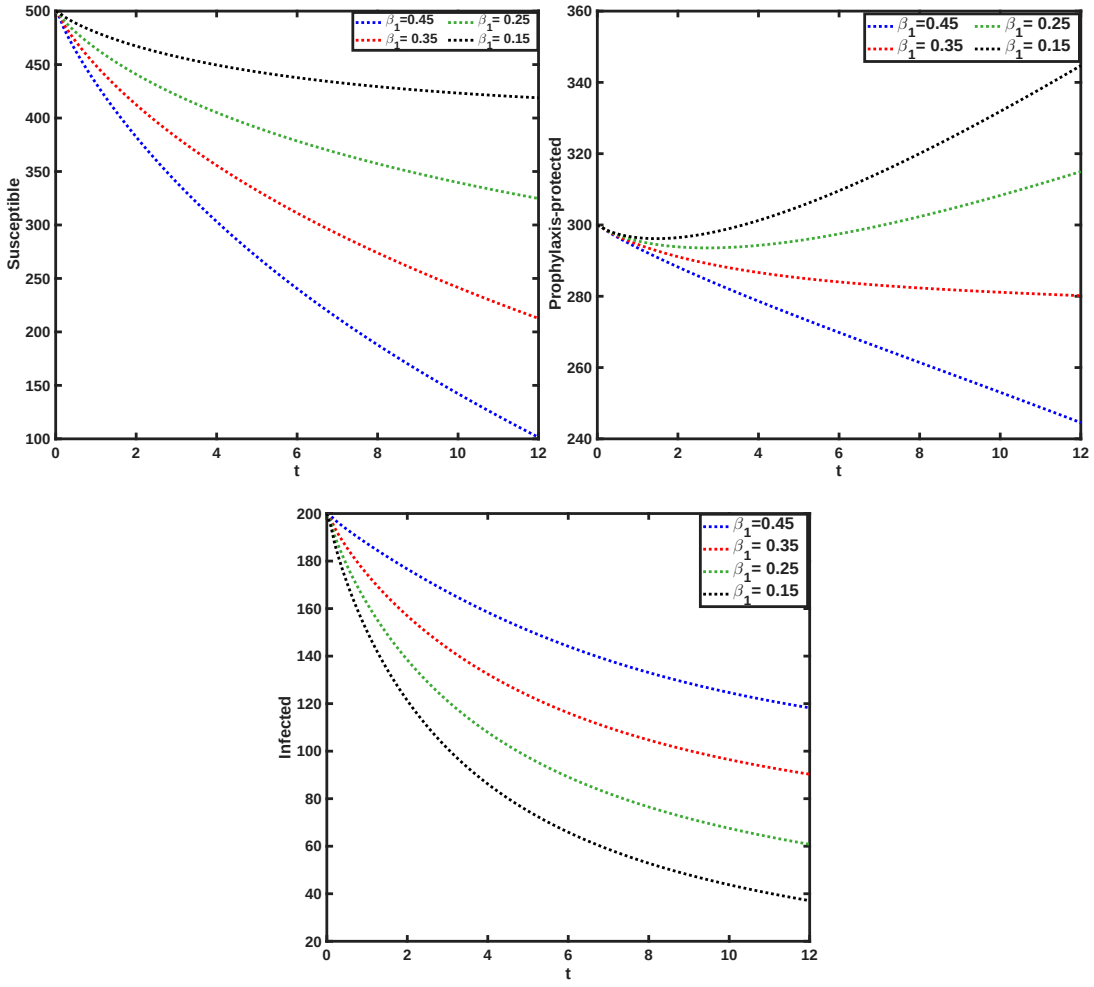


Figure 6: Effect of parameter β_1 on population dynamics.

of solutions, the derivation of the basic reproduction number, and the stability conditions of both the disease-free and endemic EPs. Furthermore, the sensitivity analysis pinpointed the parameters that most significantly affect disease transmission.

Numerical experiments show that high values of transmission rates (β_1 and β_2) result in greater population-wide infection and, subsequently, a more drastic reduction in the number of individuals who remain healthy. Hence, these findings substantiate that contact with infected individuals in the population is the main transmission factor of the disease. However, increases in the efficiency of the prophylactic pharmaceutical drug (ψ) and the fraction of newly infected susceptibles who receive prophylaxis (p) can lead to lower contagion levels and an increase in the

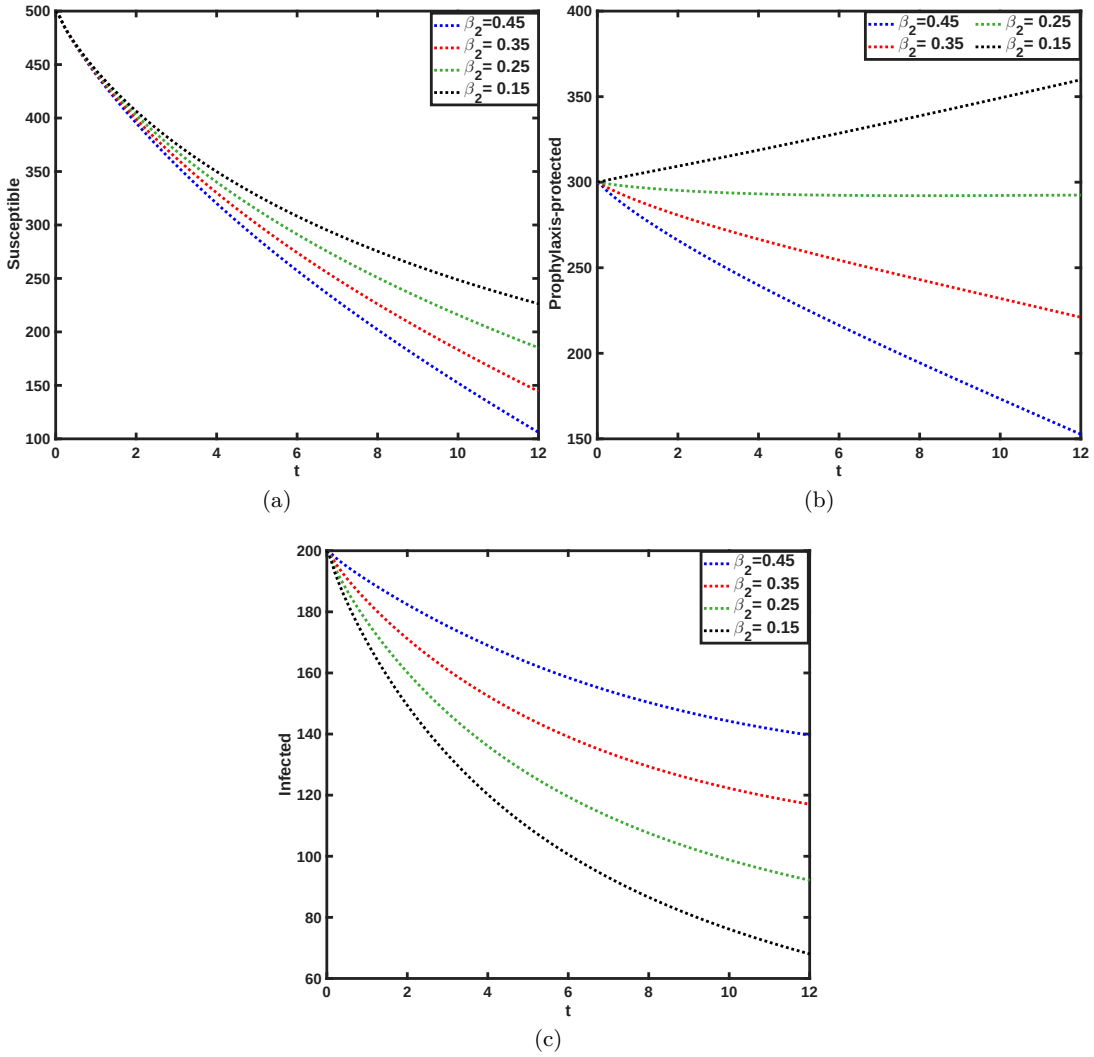


Figure 7: Effect of parameter β_2 on population dynamics.

overall protective power of the population. Moreover, switching to the fractional-order model indicated that the memory effect can influence not only the speed of disease propagation but also the disease pattern in the long run, thereby providing a more complete understanding of changes over time in an epidemic.

The findings of the present study are consistent with previous pneumonia modeling studies [26, 29, 21, 20, 54], which reported that reducing disease transmission and strengthening preventive interventions can significantly limit infection spread. Similarly, the fractional-order pneumonia study in [33] highlighted the important role of the force of infection in disease control.

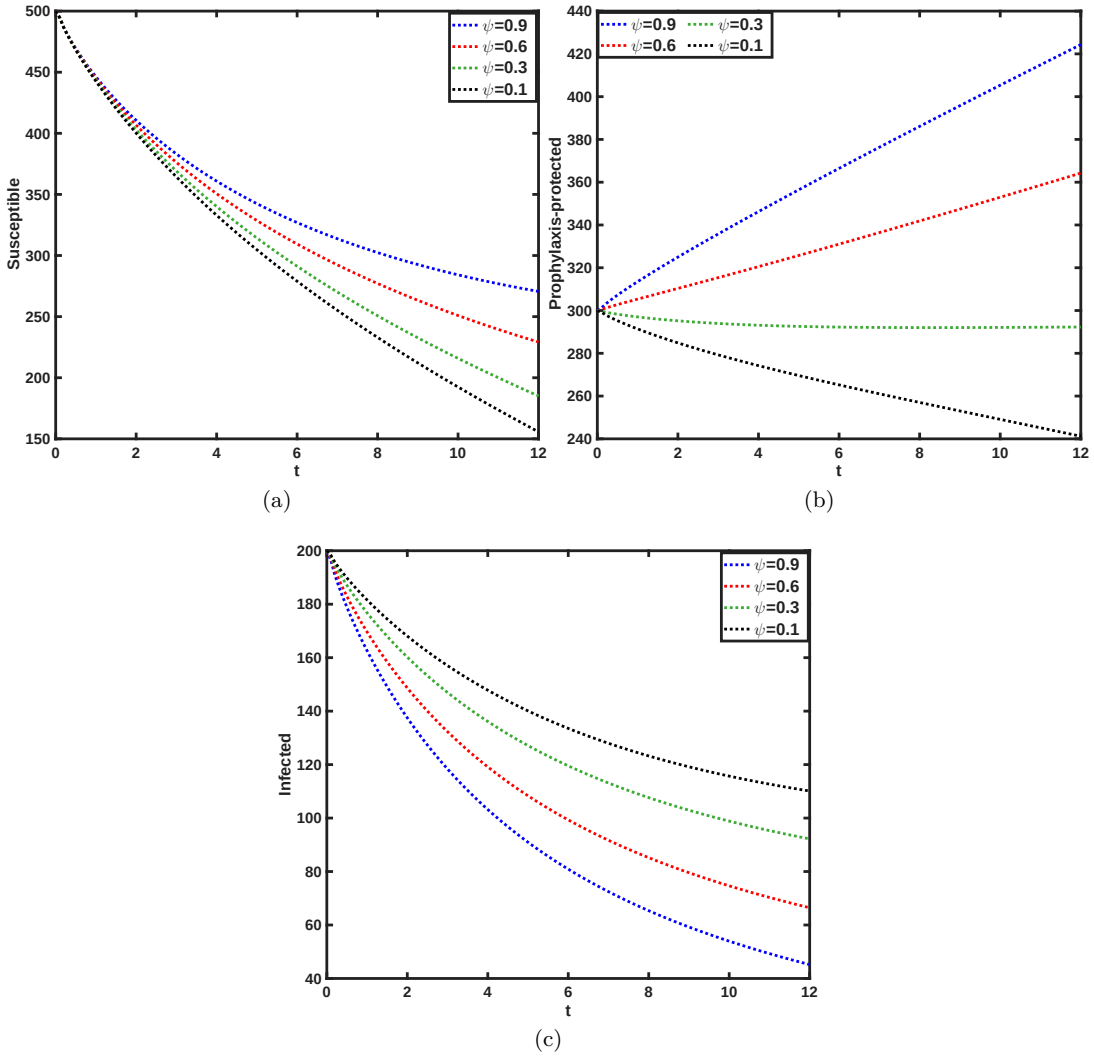


Figure 8: Effect of parameter ψ on population dynamics.

However, unlike these previous studies, the present results specifically demonstrate that increasing prophylaxis coverage and efficacy reduces the infected population in immunocompromised individuals. Moreover, the fractional derivative in the Caputo sense analysis reveals the influence of memory effects on the temporal disease dynamics, providing an additional perspective that is not captured by classical integer-order pneumonia models.

These results, from a health perspective, indicate that to achieve better disease control, a combined approach is necessary that simultaneously reduces transmission, increases the effectiveness of prophylactic drugs, and increases preventive coverage for the most vulnerable groups.

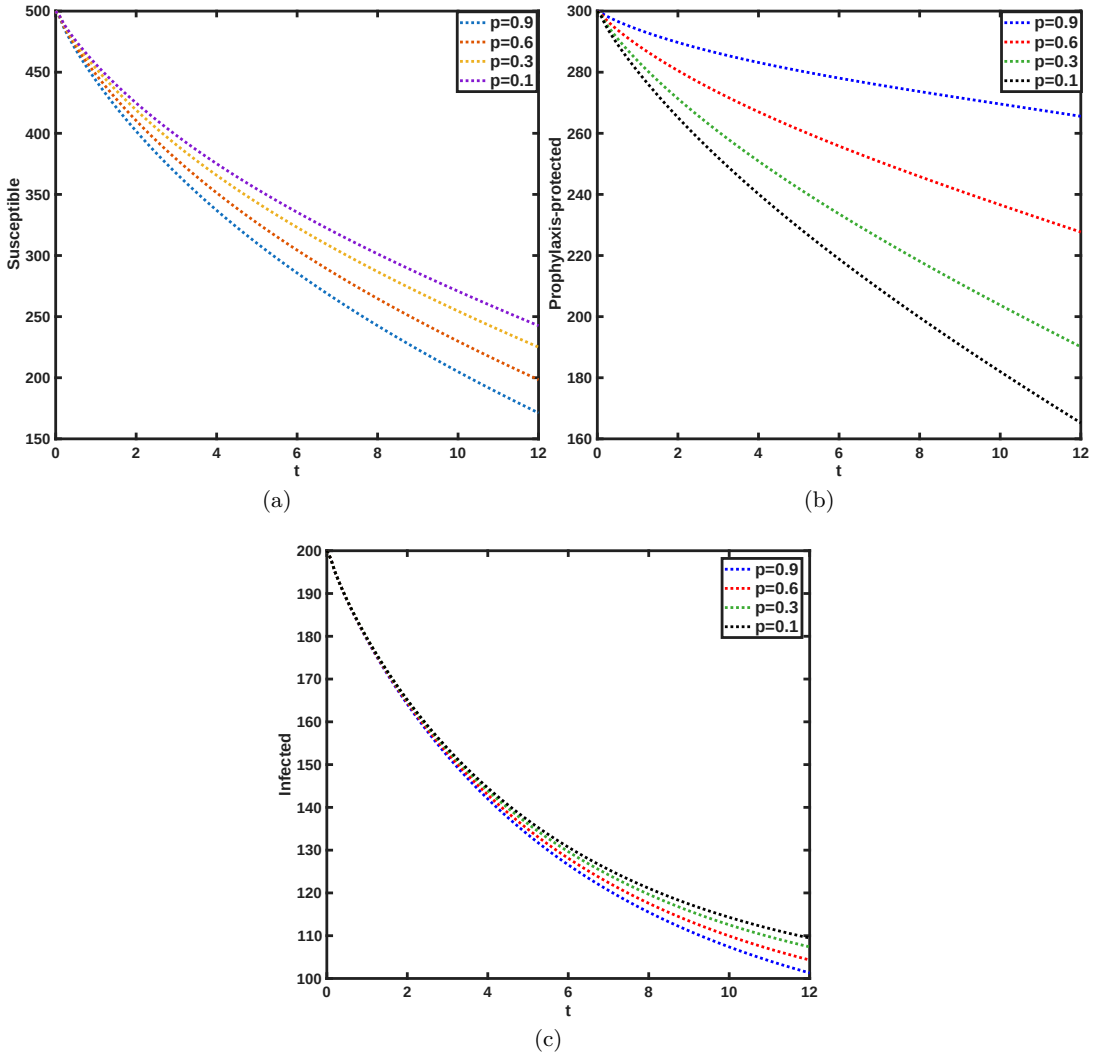


Figure 9: Effect of parameter p on population dynamics.

The study provides measurable grounds for the idea of early intervention through prevention, the use of the most efficient treatment regimens, and behaviour modifications resulting from awareness, which will lead to less contact and, eventually, lower infection rates. Finally, the study lays the groundwork for improved, more precise disease forecasting tools that would then be used to control the spread of infectious diseases among the very vulnerable groups. Additionally, it is a good instrument for public health planning that could be deliberated along with other factors, as it provides one side, the scientific evidence, of the argument. Future work can extend this model by incorporating fear effects, factors, spatial heterogeneity, and co-infection dynamics.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Almutairi, N. and El-Shahed, M., *Optimal control strategies for a mathematical model of pneumonia infection*, Computation 13(9) (2025) 204.

[2] Alqudah, M.A. and Aljahdaly, N.H., *Global stability and numerical simulation of a mathematical model of stem cells therapy of HIV-1 infection*, J. Comput. Sci. 45 (2020) 101176.

[3] Atangana, A. and Baleanu, D., *New fractional derivatives with nonlocal and non-singular kernel: Theory and application to heat transfer model*, arXiv preprint arXiv:1602.03408 (2016).

[4] Azar, M.M., Cohen, E., Ma, L., Cissé, O.H., Gan, G., Deng, Y., Belfield, K., Asch, W., Grant, M. and Gleeson, S., and Koff, A. *Genetic and epidemiologic analyses of an outbreak of Pneumocystis jirovecii pneumonia among kidney transplant recipients in the United States*, Clin. Infect. Dis. 74(4) (2022) 639–647.

[5] Batiha, I.M., Momani, S., Ouannas, A., Momani, Z. and Hadid, S.B., *Fractional-order COVID-19 pandemic outbreak: Modeling and stability analysis*, Int. J. Biomath. 15(1) (2022) 2150090.

[6] Bhatia, B., Bhatter, S., Purohit, S.D., Baleanu, D. and Suthar, D.L., *A fractional model based on Caputo derivative for tuberculosis transmission using real data from Kenya*, Sci. Rep. 16(1) (2025) 1725.

[7] Bhatia, B., Bhatter, S., Purohit, S.D. and Suthar, D., *Numerical approximation of a time-delayed fractional-order diabetes mellitus model using predictor-corrector method*, Math. Comput. Model. Dyn. Syst. 32(1) (2026) 2658274.

[8] Brown, L., Rautemaa-Richardson, R., Mengoli, C., Alanio, A., Barnes, R.A., Bretagne, S., Chen, S.C., Cordonnier, C., Donnelly, J.P. and Heinz, W.J., and Jones, B. *Polymerase chain reaction on respiratory tract specimens of immunocompromised patients to diagnose Pneumocystis pneumonia: A systematic review and meta-analysis*, Clin. Infect. Dis. 79(1) (2024) 161–168.

[9] Burton, T. and Kirk, C., *A fixed point theorem of Krasnoselskii-Schaefer type*, Math. Nachr. 189(1) (1998) 23–31.

[10] Caputo, M., *Linear models of dissipation whose Q is almost frequency independent–II*, Geophys. J. Int. 13(5) (1967) 529–539.

[11] Caputo, M. and Fabrizio, M., *A new definition of fractional derivative without singular kernel*, Prog. Fract. Differ. Appl. 1(2) (2015) 73–85.

[12] Dunbar, A., Schauwvlieghe, A., Algoe, S., Van Hellemond, J.J., Reynders, M., Vandecasteele, S., Boelens, J., Depuydt, P. and Rijnders, B., *Epidemiology of Pneumocystis jirovecii pneumonia and (non-) use of prophylaxis*, Front. Cell. Infect. Microbiol. 10 (2020) 224.

[13] Ehrhardt, M., Gašper, J. and Kilianová, S., *SIR-based mathematical modeling of infectious diseases with vaccination and waning immunity*, J. Comput. Sci. 37 (2019) 101027.

[14] Gómez-Aguilar, J. and Atangana, A., *Fractional Hunter–Saxton equation involving partial operators with bi-order in Riemann-Liouville and Liouville-Caputo sense*, Eur. Phys. J. Plus 132(2) (2017) 100.

[15] Gómez-Aguilar, J., Torres, L., Yépez-Martínez, H., Baleanu, D., Reyes, J. and Sosa, I., *Fractional Liénard type model of a pipeline within the fractional derivative without singular kernel*, Adv. Differ. Equ. 2016 (2016) 1–13.

[16] Hatzl, S., Posch, F., Scholz, L., Geiger, C., Kriegl, L., Kreuzer, P., Eller, P., Giacobbe, D.R., Bassetti, M., Hoenigl, M., and Krause, R. *Comparative efficacy and safety of treatment regimens for Pneumocystis jirovecii pneumonia in people living with HIV: A systematic review and network meta-analysis of randomized controlled trials*, Clin. Microbiol. Infect. 31(5) (2025) 713–723.

[17] Jain, H. and Sinha, A.K., *Modeling the effect of Wolbachia to control malaria transmission*, Expert Syst. Appl. 221 (2023) 119769.

[18] Jaramillo Cartagena, A., Asowata, O.E., Ng, D. and Babady, N.E., *An overview of the laboratory diagnosis of Pneumocystis jirovecii pneumonia*, J. Clin. Microbiol. 63(3) (2025) e00361–24.

[19] Jin, F., Liang, H., Chen, W.-C., Xie, J. and Wang, H.-L., *Development and validation of tools for predicting the risk of death and ICU admission of non-HIV-infected patients with Pneumocystis jirovecii pneumonia*, Front. Public Health 10 (2022) 972311.

[20] Kanyiri, C.W., Luboobi, L. and Kimathi, M., *Application of optimal control to influenza pneumonia coinfection with antiviral resistance*, Comput. Math. Methods Med. 2020(1) (2020) 5984095.

[21] Kizito, M. and Tumwiine, J., *A mathematical model of treatment and vaccination interventions of pneumococcal pneumonia infection dynamics*, J. Appl. Math. 2018(1) (2018) 2539465.

[22] Kumawat, S., Bhatter, S., Bhatia, B., Purohit, S.D., Baskonus, H.M. and Suthar, D.L., *Novel application of q-HAGTM to analyze Hilfer fractional differential equations in diabetic dynamics*, Math. Comput. Simul. 238 (2025) 136–149.

[23] Kumawat, S., Bhatter, S., Purohit, S.D., Akgül, A. and Hafez, M., *Assessment of brucellosis dynamics through fractional-order models with empirical validation*, Progr. Fract. Differ. Appl. 12(2) (2026) 341–357.

[24] Kushavaha, S.K., Soni, K. and Sinha, A.K., *A fractional-order Zika model with saturated incidence and multiple transmission routes: Analysis and cost-effective optimal control*, Int. J. Dyn. Control 14(188) (2026) 1–34.

[25] Levin, A.S.S., Costa, S., Gobara, S., Caiaffa Filho, H. and Ianhez, L., *A Pneumocystis carinii pneumonia outbreak in a renal transplant unit*, Infect. Control Hosp. Epidemiol. 21(2) (2000) 138–139.

[26] Manuputty, A.H., Lesnussa, Y.A. and Leleury, Z.A., *Stability analysis of pneumonia mathematic model with prevention using vaccines and treatment*, AIP Conf. Proc. 2360 AIP Publishing LLC, 2021.

[27] McDonald, E.G., Afshar, A., Assiri, B., Boyles, T., Hsu, J.M., Khuong, N., Prosty, C., So, M., Sohani, Z.N., Butler-Laporte, G., and Lee, T.C. *Pneumocystis jirovecii pneumonia in people living with HIV: A review*, Clin. Microbiol. Rev. 37(1) (2024) e00101–22.

[28] Odibat, Z.M. and Shawagfeh, N.T., *Generalized Taylor's formula*, Appl. Math. Comput. 186(1) (2007) 286–293.

[29] Ossaiugbo, M.I. and Okposo, N.I., *Mathematical modeling and analysis of pneumonia infection dynamics*, Sci. World J. 16(2) (2021) 73–80.

[30] Peter, O.J., Yusuf, A., Oshinubi, K., Oguntolu, F.A., Lawal, J.O., Abioye, A.I. and Ayoola, T.A., *Fractional order of pneumococcal pneumonia infection model with Caputo–Fabrizio operator*, Results Phys. 29 (2021) 104581.

[31] Reid, A.B., Chen, S.C.-A. and Worth, L.J., *Pneumocystis jirovecii pneumonia in non-HIV-infected patients: New risks and diagnostic tools*, Curr. Opin. Infect. Dis. 24(6) (2011) 534–544.

[32] Saad, K.M. and Gómez-Aguilar, J.F., *Analysis of reaction–diffusion system via a new fractional derivative with non-singular kernel*, Physica A 509 (2018) 703–716.

[33] Saber, S., Alghamdi, A.M., Ahmed, G.A. and Alshehri, K.M., *Mathematical modeling and optimal control of pneumonia disease in sheep and goats in Al-Baha region with cost-effective strategies*, AIMS Math. 7(7) (2022) 12011–12049.

[34] Samko, S.G., *Fractional integrals and derivatives: Theory and applications*, Gordon and Breach, 1993.

[35] Sauvat, L., Denis, L., Nourrisson, C., Poirier, P., Ruivard, M. and Le Guenno, G., *Pneumocystis jirovecii pneumonia in non-HIV patients: Need for a more extended prophylaxis*, Front. Med. 11 (2024) 1414092.

[36] Singh, S.P. and Chandrasekar, P.H., *Pneumocystis jiroveci pneumonia (PJP)*, Medscape (2023). Available at: https://emedicine.medscape.com/article/225976-overview. Accessed: 16 February 2026.

[37] Sinha, A.K. and Soni, K., *Fractional-order modeling of hepatitis B dynamics incorporating socio-environmental factors*, Discov. Public Health 22(1) (2025) 537.

[38] Sinha, A.K., Soni, K. and Purohit, S.D., *A novel fractional mathematical model for the impact of media awareness on hepatitis B*, Math. Open, 5 (2026).

[39] Sintunavarat, W. and Turab, A., *Mathematical analysis of an extended SEIR model of COVID-19 using the ABC-fractional operator*, Math. Comput. Simul. 198 (2022) 65–84.

[40] Solís-Pérez, J., Gómez-Aguilar, J. and Atangana, A., *Novel numerical method for solving variable-order fractional differential equations with power, exponential and Mittag-Leffler laws*, Chaos Solitons Fractals 114 (2018) 175–185.

[41] Soni, K., Kumawat, S. and Sinha, A.K., *Dynamics of Marburg virus in the presence of burial and cremation practices: A fractional approach*, Math. Comput. Sci. 6(3) (2025) 1–22.

[42] Soni, K., Kumawat, S. and Sinha, A.K., *Fractional-order model for Marburg virus transmission: Assessing the impact of awareness, burial and cremation practices*, Beni-Suef Univ. J. Basic Appl. Sci. 14(1) (2025) 109.

[43] Soni, K., Kushavaha, S.K. and Sinha, A.K., *Mathematical model of malaria transmission with prophylactic treatment using Caputo fractional operator*, Frankl. Open (2026) 100688.

[44] Soni, K., Shyamsunder and Sinha, A.K., *Memory-driven dynamics of hepatitis B transmission incorporating media awareness*, Adv. Theory Simul. 9 (2026) e70394.

[45] Soni, K. and Sinha, A.K., *Modeling and stability analysis of the transmission dynamics of monkeypox with control intervention*, Partial Differ. Equ. Appl. Math. 10 (2024) 100730.

[46] Soni, K. and Sinha, A.K., *Modeling Marburg virus control with limited hospital beds: A fractional approach*, Phys. Scr. 100(1) (2024) 015251.

[47] Soni, K. and Sinha, A.K., *Dynamics of epidemic model with conformable fractional derivative*, Nonlinear Sci. 4 (2025) 100040.

[48] Soni, K. and Sinha, A.K., *A novel fractional-order mathematical model for hepatitis B: Impact of awareness*, Int. J. Model. Simul. (2025) 1–18.

[49] Soni, K. and Sinha, A.K., *A fractional-order model for hepatitis B transmission in the framework of nonsingular and nonlocal kernel*, Palest. J. Math. 15(1) (2026) 487–516.

[50] Stavros Anevlavis, M., Kaltsas, K. and Demosthenes Bouros, M., *Prophylaxis for Pneumocystis pneumonia (PCP) in non-HIV infected patients*, Pneumon. 25(4) (2012) 348–350.

[51] Stop Pneumonia / Every Breath Counts, *The forgotten pandemic: 2.5 million annual pneumonia deaths*, (2023). Available at: https://stoppneumonia.org/latest/the-forgotten-pandemic/. Accessed: 16 February 2026.

[52] Syam, M. and Al-Refai, M., *Fractional differential equations with Atangana–Baleanu fractional derivative: Analysis and applications*, Chaos Solitons Fractals X 2 (2019) 100013.

[53] Tasaka, S. and Tokuda, H., *Pneumocystis jirovecii pneumonia in non-HIV-infected patients in the era of novel immunosuppressive therapies*, J. Infect. Chemother. 18(6) (2012) 793–806.

[54] Teklu, S.W., *Investigating the effects of intervention strategies on pneumonia and HIV/AIDS coinfection model*, Biomed. Res. Int. 2023(1) (2023) 5778209.

[55] Teklu, S.W. and Mekonnen, T.T., *HIV/AIDS-pneumonia coinfection model with treatment at each infection stage: Mathematical analysis and numerical simulation*, J. Appl. Math. 2021(1) (2021) 5444605.

[56] Teklu, S.W. and Rao, K.P., *HIV/AIDS-pneumonia codynamics model analysis with vaccination and treatment*, Comput. Math. Methods Med. 2022(1) (2022) 3105734.

[57] UNICEF Data, *Pneumonia*, UNICEF Data: Monitoring the Situation of Children and Women (2023). Available at: https://data.unicef.org/topic/child-health/pneumonia/. Accessed: 16 February 2026.

[58] Xue, T., Kong, X. and Ma, L., *Trends in the epidemiology of Pneumocystis pneumonia in immunocompromised patients without HIV infection*, J. Fungi 9(8) (2023) 812.

[59] Yadav, L., Gour, M., Purohit, S., Al-Mdallal, Q. and Suthar, D., *Generalized Euler method to study the vaccination effects on dynamics of measles infection model under non-singular kernel*, Sci. Rep. 16(1) (2026) 10429.

[60] Yiannakis, E.P. and Boswell, T.C., *Systematic review of outbreaks of Pneumocystis jirovecii pneumonia: Evidence that P. jirovecii is a transmissible organism and the implications for healthcare infection control*, J. Hosp. Infect. 93(1) (2016) 1–8.

Iranian Journal of Numerical Analysis and Optimization
Vol. 17, No. 1, 2027, pp 332–358
https://doi.org/10.22067/ijnao.2026.99592.1903
https://ijnao.um.ac.ir/

CC BY NC

Research Article

A conjugate gradient algorithm based on a combined descent scheme for two-layer neural network training and unconstrained optimization

C. Tebbal, M. Laouar, R. Ziadi*, and A.E. Mehamdia

Abstract

Conjugate gradient methods are considered among the most efficient methods for solving large-scale optimization problems. Among them, the CD (conjugate descent) variant has attracted the attention of many researchers thanks to its straightforward iterative process and low memory requirements. In the present work, we propose an improved version of the CD method to address large-scale unconstrained optimization problems, with a specific application to training two-layer neural network models. The sufficient descent condition is analyzed under the strong Wolfe line search conditions, and global convergence is established under mild assumptions.

*Corresponding author

Received 28 June 2026; revised 25 August 2026; accepted 19 September 2026

Chalabia Tebbal and Raouf Ziadi
Laboratory of Fundamental and Numerical Mathematics (LMFN), Department of Mathematics, University Setif - 1 - Ferhat Abbas, Setif, Algeria. e-mail: {chalabia.tebbal, ziadi.raouf}@gmail.com

Mounia Laouar
Laboratory of Partial Differential Equations, Department of Mathematics, Faculty of Mathematics and Computer Science, University of Batna 2, Batna, Algeria. e-mail: m.laouar@univ-batna2.dz

Abd Elhamid Mehamdia
University M'Hamed Bougara of Boumerdes, Boumerdes 35000, Algeria. e-mail: mehamdiaabdelhamid@gmail.com

How to cite this article
Tebbal, C., Laouar, M., Ziadi, R. and Mehamdia, A.E., A conjugate gradient algorithm based on a combined descent scheme for two-layer neural network training and unconstrained optimization. *Iran. J. Numer. Anal. Optim.*, 2027; 17(1): 332-358. https://doi.org/10.22067/ijnao.2026.99592.1903

The performance of the suggested algorithm is evaluated on various two-layer neural network models, with the number of weights ranging from 32 to 4080. Furthermore, to validate its effectiveness, the approach is tested on a broad set of unconstrained benchmark functions. Comparisons with recent conjugate gradient methods confirm the robustness and efficiency of the proposed method.

AMS subject classifications (2020): 90C26, 90C30, 68T07.

Keywords: Conjugate gradient method, Conjugate descent method, neural networks, two-layer neural network training model.

1 Introduction

In this paper, we consider the following unconstrained optimization problem:

$$\min_{x\in\mathbb{R}^n} f(x), \tag{P}$$

where $f:\mathbb{R}^n\to\mathbb{R}$ is a continuously differentiable function. Problem (P) is of interest in many real-world applications requiring a differentiable cost function over an unbounded search space, such as neural network training [26], machine learning [13], and image/signal processing [17, 29]. Finding solutions to these problems becomes increasingly difficult as their dimensions increase. Scientists have explored various techniques, such as Newton's method, quasi-Newton methods, and conjugate gradient (CG) methods, to find the most efficient way to solve a particular problem. CG methods have become a favored approach due to their straightforward iterative process and low memory requirements. Starting from a point $x_0\in\mathbb{R}^n$, the sequence of points $\{x_k\}_{k\in\mathbb{N}}\subset\mathbb{R}^n$ is generated by the following recursive scheme

$$x_{k+1}=x_k+\alpha_k d_k, \tag{1}$$

where $\alpha_k\in\mathbb{R}^+$ is the step-length, which is usually determined using inexact line searches that guarantee taking steps that are neither too long nor too short, such as the standard Wolfe line search

$$\begin{aligned} &f(x_k+\alpha_k d_k)-f(x_k)\le \alpha_k\delta g_k^T d_k,\\ &g(x_k+\alpha_k d_k)^T d_k\ge \sigma g_k^T d_k, \end{aligned}$$

or the strong Wolfe line search

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \alpha_k \delta g_k^T d_k, \tag{2}$$

$$\left| g(x_k + \alpha_k d_k)^T d_k \right| \leq \sigma |g_k^T d_k|, \tag{3}$$

where $g_k = g(x_k) = \nabla f(x_k)$ denotes the gradient of f at x_k and $0 < \delta < \sigma < 1$. The descent direction d_k is typically computed as

$$d_0 = -g_0, \quad d_{k+1} = -g_{k+1} + \beta_k d_k, \quad k \in \mathbb{N}^*, \tag{4}$$

where $\beta_k \in \mathbb{R}$ is a parameter that characterizes the different CG methods. Among the most famous classical CG methods are Hestenes and Stiefel (HS) [10], Fletcher and Reeves (FR) [6], Polak–Ribière–Polyak (PRP) [19, 20], conjugate descent (CD) [7], Liu and Storey (LS) [16], and Dai and Yuan (DY) [4], whose parameters β_k are given, respectively, as follows:

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{y_k^T d_k}, \quad \beta_k^{PRP} = \frac{g_{k+1}^T y_k}{\|g_k\|^2}, \quad \beta_k^{LS} = -\frac{g_{k+1}^T y_k}{g_k^T d_k},$$

$$\beta_k^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2}, \quad \beta_k^{CD} = -\frac{\|g_{k+1}\|^2}{g_k^T d_k}, \quad \beta_k^{DY} = \frac{\|g_{k+1}\|^2}{y_k^T d_k},$$

with $y_k = g_{k+1} - g_k$, and $\|\cdot\|$ denoting the Euclidean norm in $\mathbb{R}^n$.

Over the last decade, this research topic has attracted the attention of many researchers, and numerous studies have been carried out to develop new efficient CG formulas with good numerical performance and strong convergence properties. Several new CG formulas have been proposed, including the WYL [25], RMIL [21], MZ [17], NHS [27], and others. Furthermore, various hybrid and combined CG methods have been introduced; see [15, 22, 23, 24].

Although the FR, DY, and CD methods are classical, they remain competitive and robust for large-scale problems. Nevertheless, numerous enhancements have been proposed to overcome their known limitations, such as jamming, lack of sufficient descent, or poor restart strategies, and to further improve their overall performance.

Zhu, Zhang, and Wang [28] proposed two modified DY CG methods, denoted by DDY1 and DDY2, whose corresponding conjugate parameters are given by

$$\beta_k^{DDY1} = \begin{cases} \dfrac{g_{k+1}^T \left(g_{k+1} - \frac{\mu_1 (g_{k+1}^T d_k)^2}{\|g_{k+1}\| \|g_k\| \|d_k\|^2} g_k \right)}{d_k^T (g_{k+1} - g_k)} & \text{if} \quad g_{k+1}^T g_k \geq 0, \\ \dfrac{g_{k+1}^T \left(g_{k+1} + \frac{\mu_1 (g_{k+1}^T d_k)^2}{\|g_{k+1}\| \|g_k\| \|d_k\|^2} g_k \right)}{d_k^T (g_{k+1} - g_k)} & \text{otherwise,} \end{cases}$$

$$\beta_k^{DDY2} = \begin{cases} \frac{g_{k+1}^T\left(g_{k+1} - \frac{g_{k+1}^T d_k}{\|d_k\|^2} d_k\right)}{d_k^T(g_{k+1}-g_k)+\mu_2 g_{k+1}^T d_k} & \text{if} \quad g_{k+1}^T d_k \geq 0, \\ 0 & \text{otherwise,} \end{cases}$$

with $\mu_1 \in [0,1]$ and $\mu_2 > 0$. Both variants have proven their efficiency, and their convergence analysis is conducted under the strong Wolfe conditions. Inspired by the above formulas, Guefassa and Chaib [9] introduced a CG method that converges globally under the strong Wolfe conditions, with the parameter defined as

$$\beta_k^{GICY} = \begin{cases} \frac{\|g_{k+1}\|^2 - |g_{k+1}^T d_k|}{(1-\mu)\max(0, g_{k+1}^T d_k) - \mu g_k^T d_k} & \text{if} \quad |g_{k+1}^T d_k| < \|g_{k+1}\|^2, \\ \frac{\|g_{k+1}\|^2}{(1-\mu)\max(0, g_{k+1}^T d_k) - \mu g_k^T d_k} & \text{otherwise,} \end{cases} \tag{5}$$

with $\mu \in]0,1[$.

Building on these recent developments, and in order to combine good practical performance with strong convergence properties, we propose an enhanced variant of the CD (conjugate descent) method, denoted as ICD (improved conjugate descent). The new ICD parameter is nonnegative and bounded by the classical CD parameter, satisfying $0 \leq \beta_k^{ICD} \leq \beta_k^{CD}$, which ensures that the proposed method inherits the desirable theoretical properties of CD, while practically outperforming the traditional CD method.

The remainder of this paper is organized as follows: In the next section, we introduce the new ICD parameter and the corresponding algorithm. In Section 3, we provide our proposed algorithm for two-layer neural network training, accompanied by a detailed description of the addressed problem. Next, we establish the global convergence of the proposed method in Section 4. Section 5 presents the numerical performance of the proposed approach. and Section 6 gives the conclusions.

2 The proposed conjugate parameter and the corresponding algorithm

Inspired by the GICY formula (5), we propose an improved conjugate parameter scheme for the CD method, defined as follows:

$$\beta_k^{ICD} = \begin{cases} \xi_k \frac{\|g_{k+1}\|^2 - |g_{k+1}^T g_k|}{\|g_k\| \, \|d_k\|} & \text{if} \quad |g_{k+1}^T g_k| \leq \|g_{k+1}\|^2, \\ \xi_k \frac{\|g_{k+1}\|^2}{\|g_k\| \, \|d_k\|} & \text{otherwise,} \end{cases} \tag{6}$$

where $\xi_k = \frac{|g_{k+1}^T d_k|}{-g_k^T d_k}$. This new formulation operates as a switching scheme based on the value of $\|g_{k+1}\|^2 - |g_{k+1}^T g_k|$ relative to zero. Indeed, if $\|g_{k+1}\|^2 - |g_{k+1}^T g_k| > 0$, this condition ensures the non-negativity of the ICD parameter (see Lemma 2 below). Otherwise, it defaults to a modified CD parameter that is $\xi_k \frac{\|g_{k+1}\|^2}{\|g_k\| \, \|d_k\|}$. The appearance of the factor ξ_k in both formulations plays a crucial role in controlling the growth of β_k^{ICD} to guarantee that $\beta_k^{ICD} \leq \beta_k^{CD}$ (see Lemma 2 below) and to improve numerical performance. By satisfying this upper bound, the ICD method preserves the theoretical convergence properties of the CD. Consequently, the proposed method successfully combines the theoretical strengths of the traditional CD method with superior computational efficiency. Algorithm 1 below outlines the main steps of the ICD method.

Algorithm 1 The proposed ICD algorithm

Step 0. (Initialization) Choose $x_0 \in \mathbb{R}^n$, the parameters $0 < \delta < \sigma < 1$ and a tolerance $\epsilon > 0$ for the stopping criterion.
- Compute the gradient of f at x_0, let $d_0 = -g_0$, and set $k = 0$.
Step 1. If $\|g_k\| \leq \epsilon$, then stop. Otherwise, go to the next step.
Step 2. Compute a step size α_k that meets the strong Wolfe conditions (2)–(3).
Step 3. Set $x_{k+1} = x_k + \alpha_k d_k$ and put $g_{k+1} = \nabla f(x_{k+1})$.
Step 4. Compute the new parameter β_k according formula (6) and update the search direction according to relation (4). Set $k = k + 1$, and go back to **Step 1**.

3 The ICD method for two-layer neural network training

The two-layer neural network is considered one of the fundamental models from which many modern neural network architectures have evolved, and it has contributed significantly to understanding several concepts of deep learning. It is also widely used in applications such as nonlinear function approximation, regression, pattern recognition, classification, prediction, and many other fields [2, 11, 14, 18].

This model was inspired by the structure and functioning of biological neurons in the human brain, which receive signals from other neurons, process them, and ultimately make decisions. Based on this biological inspiration, a mathematical model was developed to simulate this behavior through a set of computational units called artificial neurons, which are interconnected by adjustable weights and organized into successive layers.

This model typically consists of an input layer, a hidden layer, and an output layer. It is referred to as a "two-layer neural network" because the computational layers that contain adjustable weights are the hidden layer and the output layer, while the input layer serves only as a gateway for passing data into the network without performing any computation. The hidden layer consists of a number of neurons (units), where each neuron is connected to all inputs from the previous layer. Each neuron multiplies these inputs by their corresponding weights, sums them together, and then passes the result through a nonlinear activation function. Finally, the output of the hidden layer is multiplied by the weights of the output layer to obtain the final prediction of the network. This process is known as forward propagation. It is followed by the backpropagation phase, where the error between the network's outputs and the target outputs is propagated backward from the output layer to the input layer in order to adjust the weights and minimize the overall loss function, thereby enabling the network to improve its performance iteratively. In other words, the neural network training process is carried out through these two main phases.

Activation functions are considered the mathematical engine within an artificial neuron. Their primary importance lies in introducing nonlinearity into the system, since most real-world data follow complex nonlinear relationships. Without activation functions, the neural network would remain merely a simple linear model, incapable of learning such patterns or representing these relationships efficiently. In addition, some activation functions have the ability to control the range of outputs by restricting values within specific intervals, which contributes to improving numerical stability during the training process. Moreover, the fact that activation functions possess computable derivatives makes it easier to calculate the partial derivatives of the error function with respect to each weight through the backpropagation algorithm. Among the most commonly used activation functions are the sigmoid function, defined as $Sig(t) = \frac{1}{1+e^{-t}}$, which maps inputs into the interval $(0, 1)$, the hyperbolic tangent function given by $tanh(t) = \frac{e^{t}-e^{-t}}{e^{t}+e^{-t}}$ that outputs values in the range $(-1, 1)$, and the Rectified Linear Unit (ReLU) defined as $ReLU(t) = \max(0, t)$.

In this work, we focus on training a bias-free neural network consisting of an input layer, one hidden layer, and an output layer containing two outputs, using the sigmoid activation function, as illustrated in Figure 1.

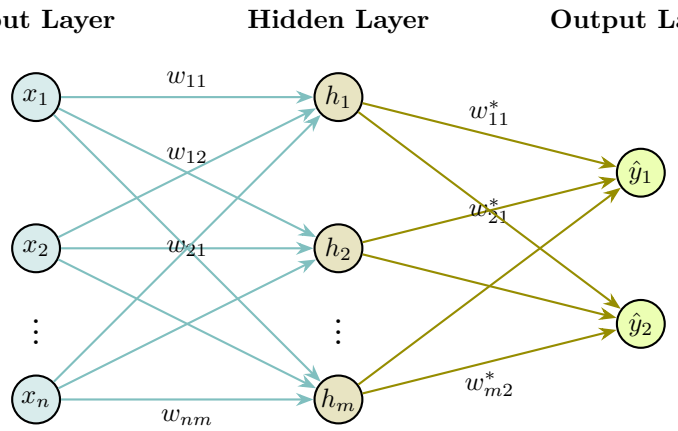


Figure 1: Architecture of a fully connected two-layer neural network.

In Figure 1, $x = (x_1, x_2, \ldots, x_n)^T$ is the input and $w = \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1m} \\ w_{21} & w_{22} & \cdots & w_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nm} \end{pmatrix}$ are the hidden layer weights, $w^* = \begin{pmatrix} w^*_{11} & w^*_{12} \\ w^*_{21} & w^*_{22} \\ \vdots & \vdots \\ w^*_{m1} & w^*_{m2} \end{pmatrix}$ are the output layer weights, while $\hat{y}_1$ and $\hat{y}_2$ denote the outputs produced by the network, and h_j denotes the output of the jth hidden neuron after applying the activation function.

During the forward propagation phase, the hidden layer values of the network are first computed by passing the weighted sums of the inputs through the sigmoid activation function

$$t_j = \sum_{i=1}^{n} w_{ij} x_i, \quad h_j = Sig(t_j), \quad j = 1 : m. \tag{7}$$

The final network outputs are then given by

$$\hat{y}_l(w, w^*) = \sum_{j=1}^{m} w^*_{jl} h_j, \quad l = 1, 2. \tag{8}$$

The loss function J represents a fundamental criterion for evaluating the performance of the model during training, as it measures the difference between the actual outputs produced by the neural network and the corresponding target values. In this model, it is defined as the sum of

squared errors for a single training sample, given by the following expression

$$J = (\hat{y}_1 - y_1)^2 + (\hat{y}_2 - y_2)^2, \tag{9}$$

where the target outputs $y_1 = 1$ and $y_2 = 0$ are used as prescribed targets for the considered training problem.

As for the backpropagation phase, the backward pass is carried out layer by layer using the chain rule with respect to the weights. This mathematical step illustrates the extent to which the weight values influence the minimization of the error function, which is represented by the partial derivatives of the error function with respect to the weights. They can be expressed as follows

$$\begin{aligned} \frac{\partial J}{\partial w_{ij}} &= \frac{\partial J}{\partial \hat{y}_1}\frac{\partial \hat{y}_1}{\partial w_{ij}} + \frac{\partial J}{\partial \hat{y}_2}\frac{\partial \hat{y}_2}{\partial w_{ij}}, \\ \frac{\partial J}{\partial w^*_{jl}} &= \frac{\partial J}{\partial \hat{y}_l}\frac{\partial \hat{y}_l}{\partial w^*_{jl}}. \end{aligned} \tag{10}$$

Accordingly, since the objective of training neural networks is to determine a set of optimal weights that minimize the loss function as much as possible, this process is essentially considered an optimization problem. Moreover, since the loss function is continuous and differentiable with respect to the weights, CG methods can be employed to train two-layer neural network models in order to compute the optimal weight values. Unlike the standard gradient descent methods, which may oscillate or converge slowly, the CG method utilizes previous gradient information to determine more efficient search directions, leading to faster convergence. Therefore, our proposed method ICD can be used to update the weights using the following iterative formula

$$W_{k+1} = W_k + \theta_k d_k, \tag{11}$$

where θ_k represents the learning rate, which is determined using line search techniques, and $W \in \mathbb{R}^{m(n+2)}$ is the column vector of the total weights, defined as

$$W = \left(V_1, \ldots, V_m, V_1^*, V_2^*\right)^T, \tag{12}$$

where

$$V_j = (w_{1j}, w_{2j}, \ldots, w_{nj}), \text{ and } V_l^* = (w_{1l}^*, w_{2l}^*, \ldots, w_{ml}^*),$$
$$\text{for } j = 1, \ldots, m, \text{ and } l = 1, 2$$

with $V_j \in \mathbb{R}^n, j = 1, \ldots, m$, and $V_l^* \in \mathbb{R}^m, l = 1, 2$ denoting, respectively, the columns of the matrices w and w^* defined above, comprising a total of $m(n+2)$ elements. Algorithm 2 below sketches the main steps for training the two-layer neural network model using the ICD method.

Algorithm 2 The ICD algorithm for two-layer neural network training

Step 0. Initialization of the network parameters: number of input neurons n, number of hidden neurons m, input vector $x \in \mathbb{R}^n$, initial weights: $w \in \mathbb{R}^{n\times m}$ and $w^* \in \mathbb{R}^{m\times 2}$, target outputs: $y_1 = 1,\ y_2 = 0$, tolerance $\epsilon > 0$ to stop the algorithm; and define the total weights vector as in (12). Set $k = 0$.
Step 1. Perform forward propagation to compute the hidden neuron values h_j, the outputs $\hat{y}_1$ and $\hat{y}_2$, and the loss function J, respectively, following formulas (7), (8), and (9).
Step 2. Calculate the gradient vector ∇J_0 according to relation (10), and set $d_0 = -\nabla J_0$.
Step 3. If $\|\nabla J_k\| \leq \epsilon$, stop the algorithm; otherwise:
- Compute the learning rate θ_k using the strong Wolfe line search conditions.
- Update the weight vector $\{W_k\}_{k\in\mathbb{N}}$ according to formula (11).
- Compute the new gradient ∇J_{k+1}, and update the search direction using the ICD formula: $d_{k+1} = -\nabla J_{k+1} + \beta_k^{ICD} d_k$.
- Set $k = k + 1$, and repeat **Step 3**.

Examples 1–3 display the models with the obtained approximate optimum weights (w, w^*).

Example 1: three inputs with one hidden layer containing four neurons.

Example 2: four inputs with one hidden layer containing five neurons.

Example 3: five inputs with one hidden layer containing five neurons.

For all experiments, the desired target outputs were fixed as 1 and 0. Figures 2, 3, and 4 display their corresponding graphical illustrations.

Example 1. As shown in Figure 2, a two-layer neural network is considered with $n = 3$ inputs and $m = 4$ hidden neurons, resulting in a total of 20 trainable weights. Algorithm 2 is applied using an initial input vector, $x = (0.17, 0.35, 0.57)^T$, and initial weights,

$$w = \begin{pmatrix} 0.1 & 1 & 0.1 & 1 \\ 0.1 & 1 & 0.1 & 1 \\ 0.1 & 1 & 0.1 & 1 \end{pmatrix},\ w^* = \begin{pmatrix} 0.1 & 1 \\ 0.1 & 1 \\ 0.1 & 1 \\ 0.1 & 1 \end{pmatrix}.$$

After training, the resulting approximate optimal weights are

$$w = \begin{pmatrix} -1.0236 & -0.4161 & -1.0236 & -0.4161 \\ -2.2133 & -1.9155 & -2.2133 & -1.9155 \\ -3.6674 & -3.7480 & -3.6674 & -3.7480 \end{pmatrix},\ w^* = \begin{pmatrix} 5.4766 & -0.3870 \\ 4.6883 & 0.3319 \\ 5.4766 & -0.3870 \\ 4.6883 & 0.3319 \end{pmatrix},$$

with $(\hat{y}_1, \hat{y}_2)^T \simeq (1, 0)$, and an error value of $J = 8.0201e - 17$.

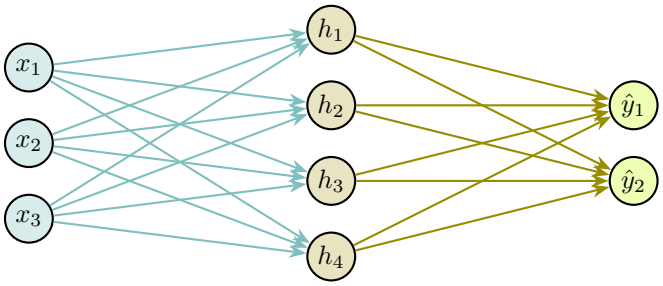


Figure 2: Neural network architecture with 3 inputs and 20 weights.

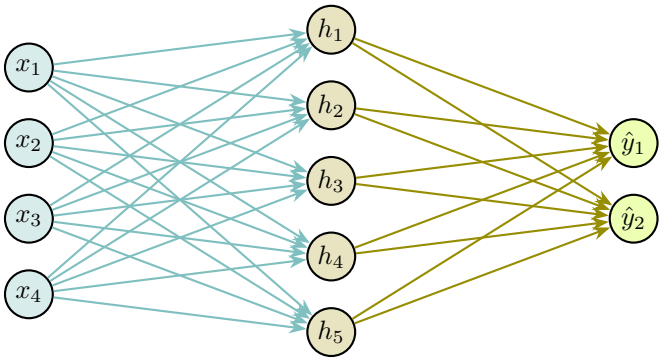


Figure 3: Neural network architecture with 4 inputs and 30 weights.

Example 2. As shown in Figure 3, a two-layer neural network is considered with $n = 4$ inputs and $m = 5$ hidden neurons, resulting in a total of 30 trainable weights. Algorithm 2 is applied using an initial input vector, $x = (0.05, 0.26, 0.66, 0.87)^T$, and initial weights,

$$w = \begin{pmatrix} 0.1 & 1 & 0.1 & 1 & 0.1 \\ 1 & 0.1 & 1 & 0.1 & 1 \\ 0.1 & 1 & 0.1 & 1 & 0.1 \\ 1 & 0.1 & 1 & 0.1 & 1 \end{pmatrix}, \; w^* = \begin{pmatrix} 0.1 & 1 \\ 0.1 & 1 \\ 0.1 & 1 \\ 0.1 & 1 \\ 0.1 & 1 \end{pmatrix}.$$

After training, the algorithm yields the approximate optimal weights:

$$w = \begin{pmatrix} -0.0635 & 0.8553 & -0.0635 & 0.8553 & -0.0635 \\ 0.1496 & -0.6527 & 0.1496 & -0.6527 & 0.1496 \\ -2.0588 & -0.9106 & -2.0588 & -0.9106 & -2.0588 \\ -1.8457 & -2.4186 & -1.8457 & -2.4186 & -1.8457 \end{pmatrix}, \; w^* = \begin{pmatrix} 4.0062 & 0.3224 \\ 3.5075 & -0.4413 \\ 4.0062 & 0.3224 \\ 3.5075 & -0.4413 \\ 4.0062 & 0.3224 \end{pmatrix}.$$

with $(\hat{y}_1, \hat{y}_2)^T \simeq (1, 0)$, and an error value of $J = 1.8947e - 16$.

Example 3. In this example, as shown in Figure 4, a two-layer neural network is considered with $n = 5$ inputs and $m = 5$ hidden neurons, resulting in a total of 35 trainable weights. Algorithm 2 is applied using an initial input vector, $x = (0.2, 0.4, 0.6, 0.8, 0.15)^T$, and initial weights,

$$w = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 & 2 \\ 3 & 3 & 3 & 3 & 3 \\ 4 & 4 & 4 & 4 & 4 \\ 5 & 5 & 5 & 5 & 5 \end{pmatrix}, \; w^* = \begin{pmatrix} 0.5 & 1 \\ 1 & 0.5 \\ 0.5 & 1 \\ 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}.$$

After training, the algorithm yields the approximate optimal weights:

$$w = \begin{pmatrix} 0.3224 & 0.3230 & 0.3224 & 0.3230 & 0.3224 \\ 0.6448 & 0.6459 & 0.6448 & 0.6459 & 0.6448 \\ 0.9672 & 0.9689 & 0.9672 & 0.9689 & 0.9672 \\ 1.2897 & 1.2919 & 1.2897 & 1.2919 & 1.2897 \\ 4.4918 & 4.4922 & 4.4918 & 4.4922 & 4.4918 \end{pmatrix}, \; w^* = \begin{pmatrix} 0.0126 & 0.1967 \\ 0.5178 & -0.2949 \\ 0.0126 & 0.1967 \\ 0.5178 & -0.2949 \\ 0.0126 & 0.1967 \end{pmatrix},$$

with $(\hat{y}_1, \hat{y}_2)^T \simeq (1, 0)$, and an error value of $J = 1.2360e - 18$.

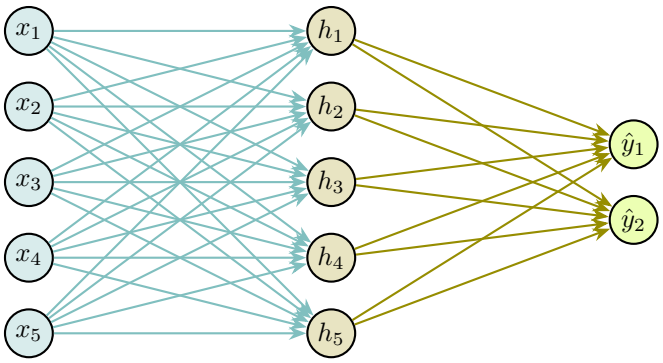


Figure 4: Neural network architecture with 5 inputs and 35 weights.

4 Convergence analysis of the ICD method

Before studying the convergence of the ICD method, we first establish in the following lemma the descent property of the generated directions $\{d_k\}_{k\in\mathbb{N}}$.

4.1 The sufficient descent condition

Lemma 1. Let $\{\beta_k\}_{k\in\mathbb{N}}$, $\{g_k\}_{k\in\mathbb{N}}$ and $\{d_k\}_{k\in\mathbb{N}}$ be generated by the ICD algorithm using the strong Wolfe conditions (2)–(3). Then the search directions $\{d_k\}_{k\in\mathbb{N}}$ satisfy the sufficient descent condition

$$g_k^T d_k \leq -c\|g_k\|^2, \quad \text{with } c > 0. \tag{13}$$

Proof. The proof proceeds by induction. It is clear that the property holds for $k = 0$, since $g_0 = -d_0$. Assume that this relation holds for k. Then, using the second strong Wolfe condition (3), we get

$$0 \leq \xi_k \leq \sigma, \tag{14}$$

with $\xi_k = \frac{|g_{k+1}^T d_k|}{-g_k^T d_k}$. By the definition of β_k^{ICD}, from relation (14) it follows that $\beta_k^{ICD} \leq \sigma \frac{\|g_{k+1}\|^2}{\|g_k\|\,\|d_k\|}$. Indeed,
- Case (i): If $|g_{k+1}^T g_k| \leq \|g_{k+1}\|^2$, then

$$\beta_k^{ICD} = \xi_k \frac{\|g_{k+1}\|^2 - |g_{k+1}^T g_k|}{\|g_k\|\|d_k\|} \leq \sigma \frac{\|g_{k+1}\|^2}{\|g_k\|\|d_k\|}.$$

- Case (ii): If $|g_{k+1}^T g_k| > \|g_{k+1}\|^2$, then

$$\beta_k^{ICD} = \xi_k \frac{\|g_{k+1}\|^2}{\|g_k\|\,\|d_k\|} \leq \sigma \frac{\|g_{k+1}\|^2}{\|g_k\|\,\|d_k\|}.$$

Therefore, using the recursive scheme (4), and from the second strong Wolfe condition (3), we get

$$g_{k+1}^T d_{k+1} \leq -\|g_{k+1}\|^2 + \beta_k^{ICD} |g_{k+1}^T d_k| \tag{15}$$

$$\leq -(1-\sigma^2)\|g_{k+1}\|^2. \tag{16}$$

□

The following lemma shows that the ICD method preserves the theoretical properties of the CD method, a feature that is essential for ensuring global convergence.

Lemma 2. As defined in (6), the proposed CG parameter β_k^{ICD} is bounded between 0 and β_k^{CD}; namely

$$0 \leq \beta_k^{ICD} \leq \beta_k^{\mathrm{CD}}, \quad \text{for all } k \in \mathbb{N}.$$

Proof. Following the definition of β_k^{ICD}, we show that $0 \leq \beta_k^{ICD} \leq \beta_k^{\mathrm{CD}}$ in both cases.
- Case (i): If $|g_{k+1}^T g_k| \leq \|g_{k+1}\|^2$, by the positivity of ξ_k ($k \in \mathbb{N}$), from (14) it follows that

$$\beta_k^{ICD} = \xi_k \frac{\|g_{k+1}\|^2 - |g_{k+1}^T g_k|}{\|g_k\| \, \|d_k\|} \geq 0,$$

on the other hand, using the Cauchy-Schwarz inequality, we get

$$\beta_k^{ICD} \leq \sigma \frac{\|g_{k+1}\|^2}{\|g_k\| \|d_k\|} \leq \sigma \frac{\|g_{k+1}\|^2}{|g_k^T d_k|} \leq \beta_k^{CD}.$$

- Case (ii): If $|g_{k+1}^T g_k| > \|g_{k+1}\|^2$, then

$$0 \leq \beta_k^{ICD} = \xi_k \frac{\|g_{k+1}\|^2}{\|g_k\| \, \|d_k\|} \leq \sigma \frac{\|g_{k+1}\|^2}{|g_k^T d_k|} \leq \beta_k^{CD}.$$

Therefore, the proof is valid for both cases.

□

4.2 Global convergence

For the global convergence, we require the following assumptions.

Assumption 1. The level set $\Omega = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$ is bounded.

Assumption 2. In a closed neighborhood $\mathcal{N}$ of Ω, the objective function f is continuously differentiable and its gradient g is Lipschitz continuous; namely, there exists a positive constant L such that

$$\|g(x) - g(y)\| \leq L\|x - y\|, \qquad \text{for all } x, y \in \mathcal{N}. \tag{17}$$

These assumptions imply that there exists a positive constant $\Gamma \geq 0$, such that

$$\|g(x)\| \leq \Gamma, \qquad \text{for all } x \in \mathcal{N}. \tag{18}$$

To establish the global convergence of the ICD method, we make use of a lemma by Dai et al. [3].

Lemma 3. Consider any iterative method of the form (1) and (4), where d_k is a descent direction and α_k meets the strong Wolfe conditions. Then, under the above assumptions, if

$$\sum_{k\geq 0} \frac{1}{\|d_k\|^2} = +\infty,$$

then

$$\liminf_{k\to\infty} \|g_k\| = 0$$

We also need the following lemma.

Lemma 4. Under Assumption 2, the sequence of step sizes $\{\alpha_k\}_{k\in\mathbb{N}}$ is strictly positive, with

$$\alpha_k \geq \frac{(\sigma-1)}{L}\frac{g_k^T d_k}{\|d_k\|^2}, \quad k \in \mathbb{N}. \tag{19}$$

Proof. From the second strong Wolfe condition (3), we have

$$(g_{k+1} - g_k)^T d_k \geq (\sigma - 1) g_k^T d_k.$$

Under Assumption 2, using the Cauchy–Schwarz inequality, we get

$$L\|x_{k+1} - x_k\| \, \|d_k\| \geq (\sigma - 1) g_k^T d_k,$$

therefore,

$$\alpha_k \geq \frac{\sigma - 1}{L}\frac{g_k^T d_k}{\|d_k\|^2}.$$

and the proof is complete. □

Note that from relation (16), $\alpha_k \geq (\sigma-1)^2(\sigma+1)\frac{\|g_k\|^2}{L\|d_k\|^2}$, for all $k \in \mathbb{N}$. It is clear that, if x_k is not the optimal solution, then $\|g_k\| > \epsilon$; hence, according to the strong Wolfe conditions (3) $(0 < \sigma < 1)$, it follows that the step size α_k is strictly positive for all $k \in \mathbb{N}$; that is, there exists a constant $\tilde{m} > 0$ such that $\alpha_k \geq \tilde{m}$, for all $k \geq 0$.

We now establish our main results, which concern the global convergence of the ICD algorithm.

Theorem 1. Let the sequences $\{g_k\}_{k\geq 0}$, $\{d_k\}_{k\geq 0}$ be generated by the ICD algorithm. Under the above considerations, the ICD algorithm converges in the sense that

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{20}$$

Proof. Assume, for the sake of contradiction, that the convergence does not hold. Then, there exists $\gamma > 0$ such that

$$\|g_k\| \geq \gamma, \quad \text{for all } k \in \mathbb{N}. \tag{21}$$

From Lemma 2 and relations (18) and (21), it holds that

$$\begin{aligned} \beta_k^{ICD} &\leq \frac{\|g_{k+1}\|^2}{-g_k^T d_k} \\ &\leq \frac{\Gamma^2}{c\gamma^2}. \end{aligned}$$

Let $M = \frac{\Gamma^2}{c\gamma^2}$. Since β_k^{ICD} is positive, from the last inequality and relations (4), (18), and (19), it follows that

$$\begin{aligned} \|d_{k+1}\| &\leq \|g_{k+1}\| + \beta_k^{ICD}\|d_k\| \\ &\leq \Gamma + M\frac{D}{\tilde{m}}, \end{aligned}$$

where

$$D = \max\{\|x - y\| : x, y \in \mathcal{N}\}.$$

This implies that

$$\sum_{k\geq 0} \frac{1}{\|d_{k+1}\|^2} = +\infty. \tag{22}$$

By applying Lemma 3, (20) holds. This contradicts (21), thus proving (20). □

5 Numerical experiments

All the numerical experiments are conducted in MATLAB R2024b and executed on a PC with an Intel(R) Core(TM) i5-8350U CPU @ 1.70 GHz, running the Windows 11 operating system. In all the test problems, all the compared CG methods are implemented under the strong Wolfe

line search conditions with $\delta = 10^{-4}$, $\sigma = 10^{-3}$. To examine the performance of the proposed method, we have compared it with five CG methods, which are CD [7], GICY [9] (with $\mu = 0.5$), IFR [12], PRP$^+$ [8], and DY [4].

5.1 Statistical analysis of two-layer neural network training

To evaluate the numerical performance and robustness of the proposed ICD algorithm, 10 test problems were considered with varying layer sizes $n = 6 - 100$ and $m = 4 - 40$, yielding a total number of weights between 32 and 4080. For each problem, we executed each algorithm for 10 independent runs, using different random seeds (x, w, w^*). Thus, a total of 100 runs were conducted for the considered test problems. The stopping criterion was set to $\|\nabla f(x)\| < 10^{-6}$ within 2×10^3 iterations. The performance of each algorithm is presented in Table 1, where we report the mean number of iterations and CPU time, the mean and standard deviation of the error, and the success rate of each algorithm over the set of implementations. The best results are highlighted in bold.

The results reported in Table 1 show that the ICD method achieves the lowest error on most runs while requiring the fewest iterations and the lowest CPU time across the ten test problems. To further support this observation, a statistical comparison was conducted using the p-values from the corresponding Wilcoxon rank-sum tests (at a 5% level of significance), comparing the proposed ICD algorithm with the CD, GICY, IFR, PRP$^+$, and DY algorithms. The statistical comparison is reported in Tables 2 and 3, where R^+ denotes the sum of runs in which ICD outperforms the corresponding competing algorithm, and R^- is the sum of runs with the opposite outcome. The score Sc indicates the outcome of the statistical comparison: $(+)$ denotes that the ICD algorithm exhibits better performance than the competing algorithm; $(-)$ denotes that the proposed algorithm exhibits inferior performance; and $(=)$ indicates no significant difference at the 5% significance level.

As shown in Table 3, ICD outperformed CD, GICY, IFR, and PRP$^+$ in 81%, 82%, 81%, and 85% of the 100 runs, respectively. Based on statistical comparisons, ICD is significantly better than CD, GICY, IFR, and PRP$^+$ in 4, 3, 6, and 5 out of the 10 test problems, respectively, with no significant difference in the remaining problems. Compared with DY, ICD achieves similar overall performance: There is no significant difference in 8 problems, while ICD outperforms DY in Problem 9 and performs worse in Problem 10, with a slight advantage to ICD in 53% of the runs.

Table 1: Numerical comparison for training two-layer neural networks over ten independent runs.

Test Problem	Method	mean_Iter	mean/std_Error	mean_CPU(s)	Success rate
Problem 1: $n = 6, m = 4$ (32 weights)	ICD	210.1	4.34e-06/1.37e-05	0.0513	90%
	CD	500.3	7.45e-08/2.35e-07	0.1104	80%
	GICY	437.8	3.49e-12/8.94e-12	0.1090	100%
	IFR	803.1	2.82e-07/8.86e-07	0.1814	60%
	PRP$^+$	211.2	4.34e-06/1.37e-05	0.0491	90%
	DY	**5.6**	**6.02e-14/9.02e-14**	**0.0006**	100%
Problem 2: $n = 6, m = 10$ (80 weights)	ICD	**36.7**	**5.44e-15/1.76e-15**	**0.0117**	100%
	CD	1023.8	1.40e-07/3.65e-07	0.2403	50%
	GICY	755.2	1.30e-11/2.64e-11	0.1886	90%
	IFR	620.7	2.06e-08/6.19e-08	0.1541	60%
	PRP$^+$	90.2	5.54e-15/1.75e-15	0.0204	100%
	DY	94.2	2.85e-14/4.68e-14	0.0240	100%
Problem 3: $n = 15, m = 10$ (170 weights)	ICD	**54.3**	**2.08e-14/4.46e-14**	**0.0127**	100%
	CD	966.4	8.66e-10/2.16e-09	0.2414	50%
	GICY	880.3	2.04e-06/6.44e-06	0.2359	60%
	IFR	1306.2	1.49e-07/2.37e-07	0.3533	10%
	PRP$^+$	420.6	7.78e-08/1.64e-07	0.1078	80%
	DY	423.8	7.76e-10/1.64e-09	0.1119	80%
Problem 4: $n = 20, m = 10$ (220 weights)	ICD	**81.8**	**6.39e-14/9.56e-14**	**0.0200**	100%
	CD	710.4	3.39e-05/1.07e-04	0.1789	40%
	GICY	801.8	6.26e-09/1.98e-08	0.2170	90%
	IFR	1601.8	3.13e-03/9.90e-03	0.4222	20%
	PRP$^+$	221.4	3.74e-08/1.18e-07	0.0569	90%
	DY	225.0	3.73e-08/1.18e-07	0.0591	90%
Problem 5: $n = 40, m = 10$ (420 weights)	ICD	20.3	2.64e-14/2.70e-14	0.0051	100%
	CD	1184.5	8.58e-04/1.79e-03	0.3303	50%
	GICY	975.6	9.28e-10/2.76e-09	0.2791	80%
	IFR	1205.8	3.92e-08/8.35e-08	0.3436	40%
	PRP$^+$	**17.1**	2.00e-14/2.12e-14	**0.0030**	100%
	DY	51.9	**1.83e-14/3.08e-14**	0.0130	100%
Problem 6: $n = 46, m = 10$ (480 weights)	ICD	**84.5**	**9.84e-14/1.38e-13**	**0.0229**	100%
	CD	849.9	5.84e-09/1.18e-08	0.2260	50%
	GICY	1160.3	2.76e-08/8.67e-08	0.3409	80%
	IFR	1411.4	3.20e-03/1.01e-02	0.4037	30%
	PRP$^+$	213.7	3.07e-08/9.70e-08	0.0584	90%
	DY	216.7	3.08e-08/9.73e-08	0.0620	90%
Problem 7: $n = 50, m = 12$ (624 weights)	ICD	**49.6**	**7.03e-13/1.98e-12**	**0.0132**	100%
	CD	1115.1	1.23e-06/3.87e-06	0.3229	70%
	GICY	917.8	2.30e-10/7.25e-10	0.2874	80%
	IFR	1407.1	5.09e-08/7.08e-08	0.4403	30%
	PRP$^+$	223.5	4.48e-08/1.42e-07	0.0644	90%
	DY	226.0	4.44e-08/1.40e-07	0.0686	90%
Problem 8: $n = 80, m = 10$ (820 weights)	ICD	410.4	2.19e-06/6.78e-06	0.1313	90%
	CD	1305.2	3.43e-04/1.01e-03	0.3939	50%
	GICY	1384.8	1.68e-05/5.30e-05	0.4426	70%
	IFR	1429.6	3.28e-06/1.01e-05	0.4417	30%
	PRP$^+$	**212.6**	3.68e-08/1.17e-07	**0.0642**	90%
	DY	271.7	**8.91e-13/1.72e-12**	0.0843	80%
Problem 9: $n = 100, m = 20$ (2040 weights)	ICD	**25.1**	**4.38e-15/3.08e-15**	**0.0095**	100%
	CD	1002.7	1.62e-05/3.44e-05	0.4218	60%
	GICY	610.3	1.32e-11/4.13e-11	0.2777	90%
	IFR	1566.8	2.41e-07/3.68e-07	0.6842	20%
	PRP$^+$	218.5	2.64e-08/8.34e-08	0.0929	90%
	DY	61.5	5.75e-12/1.81e-11	0.0241	80%
Problem 10: $n = 100, m = 40$ (4080 weights)	ICD	449.6	8.93e-08/1.88e-07	0.6244	90%
	CD	689.5	1.26e-09/3.33e-09	0.8998	50%
	GICY	379.2	9.45e-10/2.82e-09	0.5233	70%
	IFR	1616.3	6.64e-07/6.44e-07	2.2095	10%
	PRP$^+$	445.7	8.92e-08/1.88e-07	0.6166	80%
	DY	**16.6**	**5.69e-15/9.12e-15**	**0.0158**	100%

Table 2: Pairwise comparison, ICD versus other algorithms.

ICD vs ...	ICD vs CD				ICD vs GICY				ICD vs IFR			
Problems	R^+	R^-	p-value	Sc	R^+	R^-	p-value	Sc	R^+	R^-	p-value	Sc
Problem 1	4	6	0.9219	=	4	6	0.6250	=	4	6	0.9219	=
Problem 2	9	1	0.0039	+	10	0	0.0020	+	7	3	0.0273	+
Problem 3	8	2	0.4316	=	8	2	0.3223	=	8	2	0.2324	=
Problem 4	7	3	0.1934	=	8	2	0.1055	=	8	2	0.0488	+
Problem 5	9	1	0.0059	+	9	1	0.0098	+	8	2	0.0098	+
Problem 6	9	1	0.0840	=	9	1	0.0840	=	10	0	0.0020	+
Problem 7	8	2	0.0840	=	9	1	0.0840	=	9	1	0.0840	=
Problem 8	10	0	0.0020	+	8	2	0.3223	=	8	2	0.3223	=
Problem 9	9	1	0.0039	+	9	1	0.0039	+	10	0	0.0020	+
Problem 10	8	2	0.4316	=	8	2	0.4316	=	9	1	0.0273	+
ICD vs ...	ICD vs PRP^+				ICD vs DY							
Problems	R^+	R^-	p-value	Sc	R^+	R^-	p-value	Sc				
Problem 1	4	6	0.8457	=	4	6	0.4316	=				
Problem 2	10	0	0.0020	+	6	4	0.7695	=				
Problem 3	10	0	0.0020	+	7	3	0.4316	=				
Problem 4	10	0	0.0020	+	6	4	1.0000	=				
Problem 5	9	1	0.0840	=	3	7	0.4922	=				
Problem 6	8	2	0.4316	=	7	3	0.4316	=				
Problem 7	10	0	0.0020	+	6	4	0.5566	=				
Problem 8	6	4	0.7695	=	5	5	0.9219	=				
Problem 9	10	0	0.0020	+	7	3	0.0273	+				
Problem 10	8	2	0.4316	=	2	8	0.0020	-				

Table 3: Summary of statistical comparisons for 10 test problems.

Algorithms	Total R^+	Total R^-	Total (+)	Total (−)	Total (=)
ICD vs CD	81	19	4	0	6
ICD vs GICY	82	18	3	0	7
ICD vs IFR	81	19	6	0	4
ICD vs PRP^+	85	15	5	0	5
ICD vs DY	53	47	1	1	8

Overall, the results demonstrate that the ICD algorithm provides competitive performance across the considered two-layer neural network training problems.

5.2 Performance on typical benchmark functions

To validate the efficiency and robustness of the proposed ICD method, we evaluated its performance against CD [7], GICY [9], and IFR [12] across 50 test functions chosen from [1], with dimensions ranging from 2 to 50000, yielding a total of 200 test problems. The same initial starting points as suggested in [1] were used for each test problem. Each implementation is considered successful if a point x_k where $\|g_k\|_\infty \leq 10^{-6}$ is reached within 2000 iterations, with CPU-time less than 500 seconds; otherwise, the implementation is considered a failure. The detailed numerical results are reported in Tables 4, 5, and 6, listing the number of iterations (Iter), CPU time (CPU), function evaluations (NF), and gradient evaluations (NG).

Figures 5–8 illustrate the performance profiles of the compared methods with respect to the number of iterations, CPU time, the number of function evaluations, and the number of gradient evaluations. The test functions utilized are listed in Tables 4-6, and the comparative analysis is conducted using the performance profiles of Dolan and Moré [5]. The top curve in the graphs corresponds to the method that achieves the highest success rate across the test problems within the factor t, as described in [5]. Let Q denote the set of solvers, and let P be the set of test problems, with $n_q = |Q|$ and $n_p = |P|$ representing their respective cardinalities. For each problem $p \in P$ and solver $q \in Q$, let $\tau_{p,q}$ denote the performance metric (e.g., number of function or gradient evaluations) required by a solver q to solve problem p. The performance ratio is defined as

$$r_{p,q} = \frac{\tau_{p,q}}{\min\{\tau_{p,i} : i \in Q\}}.$$

Assume there exists a parameter r_M such that $r_{p,q} \leq r_M$ for all $p \in P$ and $q \in Q$, where $r_{p,q} = r_M$ if and only if a solver q fails to solve problem p. The global assessment of solver performance is provided by the performance profile function $F_q : [1, \infty) \to [0, 1]$, defined as

$$F_q(t) = \frac{1}{n_p}\text{card}\{p \in P : r_{p,q} \leq t\},$$

where card$\{\cdot\}$ denotes the cardinality of the set. The function $F_q(t)$ represents the cumulative distribution of the performance ratio.

Table 4: Numerical results of ICD, CD, GICY and IFR.

Test problem	Initial point	Dim	ICD	CD	GICY	IFR
			Iter/CPU/NF/NG	Iter/CPU/NF/NG	Iter/CPU/NF/NG	Iter/CPU/NF/NG
Almost perturbed quadratic	(0.5,...,0.5)	2	5 /0.0211/120/66	5/0.0223/120/66	12/0.0427/320/173	12/0.0570/320/173
		4	9/0.0322/238/129	9/0.0399/232/126	29/0.0245/838/449	29/0.0160/838/449
		6	10/0.0021/310/167	11/0.0034/312/168	44/0.0070/1392/741	44/0.0000/1392/741
		10	14/0.0093/472/254	17/0.0121/490/263	72/0.0080/2280/1213	72/0.0090/2280/1213
Diagonal 2	(1/1,1/2,...,1/n)	2	9/0.0388/196/108	9/0.0508/196/108	13/0.0591/326/177	13/0.0610/326/177
		10	17/0.0212/530/287	21/0.0274/532/288	60/0.0255/1600/861	60/0.0310/1694/908
		100	60/0.0501/1670/896	60/0.0509/1582/852	505/0.4378/14640/7826	499/0.3360/15086/8043
		200	55/0.0916/2168/1161	76/0.1913/2220/1187	956/3.8177/29564/15739	968/2.1470/31464/16701
Diagonal 1	(1/n,1/n,...,1/n)	2	9/0.0494/244/132	9/0.0390/244/132	12/0.0560/304/165	12/0.0460/304/165
		10	23/0.0202/852/450	23/0.0199/816/432	60/0.0331/2454/1288	60/0.0290/2450/1286
		100	144/0.1678/7392/3841	199/0.2778/10746/5573	Fail	1525/2.4870/89482/46267
		200	262/0.5319/19860/10993	384/0.9705/21856/11313	Fail	Fail
Diagonal 3	(1,...,1)	2	27/0.0323/1004/536	33/0.0589/1080/574	15/0.0502/436/234	15/0.0940/436/234
		10	26/0.0211/912/483	26/0.0207/874/464	84/0.0461/3176/1673	85/0.0250/3306/1739
		100	111/0.1182/5650/2937	159/0.2242/8698/4509	Fail	582/0.3830/27988/14577
		1000	Fail	Fail	Fail	Fail
Diagonal 4	(1,...,1)	100	7/0.0350/254/137	9/0.0434/262/141	24/0.0717/732/391	26/0.0450/806/430
		500	7/0.0187/254/137	9/0.0210/262/141	26/0.0401/796/425	26/0.0480/812/433
		20000	9/0.4114/290/156	10/0.5350/302/162	28/1.7161/866/462	28/1.5550/872/465
		30000	9/0.6396/290/156	10/0.7848/302/162	28/2.2918/872/465	30/2.3380/922/492
Diagonal 5	(1.1,...,1.1)	2	3/0.0275/50/29	3/0.0280/50/29	3/0.0321/50/29	3/0.0320/50/29
		10	3/0.0054/50/29	3/0.0064/50/29	3/0.0054/50/29	3/0.0110/50/29
		50	4/0.0285/112/61	4/0.0216/112/61	4/0.0304/112/61	4/0.0260/112/61
		100	4/0.0040/112/61	4/0.0043/112/61	4/0.0035/112/61	4/0.0060/112/61
Diagonal 7	(1,...,1)	500	4/0.0468/178/95	3/0.0205/54/31	11/0.1201/550/287	6/0.0640/240/127
		1500	5/0.1939/674/351	3/0.0199/54/31	12/0.1430/612/319	5/0.0490/178/95
		3000	4/0.2494/1532/829	3/0.0405/54/31	15/0.3379/798/415	22/0.5730/1232/639
		5000	4/2.2016/3144/1661	3/0.0443/54/31	14/0.5727/736/383	40/2.0180/2336/1209
Cosine	(1,...,1)	2	3/0.0295/94/51	3/0.0269/94/51	3/0.0304/94/51	3/0.0350/94/51
		4	45/0.0586/2664/1378	178/0.1283/10914/5636	15/0.0092/684/358	1056/0.1760/62458/32286
		10	77/0.102/3650/2215	97/0.1145/5890/3043	Fail	259/0.0820/15878/8199
		50	Fail	Fail	Fail	15/0.0110/684/358
Diagonal 8	(1,...,1)	200	4/0.0266/118/64	4/0.0336/118/64	4/0.0404/118/64	4/0.0210/118/64
		1000	4/0.0395/118/64	7/0.0760/304/160	12/0.1199/614/320	4/0.0400/118/64
		1500	4/0.0314/118/64	4/0.0336/118/64	7/0.0734/304/160	6/0.0470/242/128
		2000	4/0.0488/118/64	4/0.0386/118/64	7/0.1026/304/160	6/0.0800/242/128
Diagonal 9	(1,...,1)	4	39/0.0609/1918/999	32/0.0572/1550/808	51/0.0289/2520/1312	132/0.0328/7316/3791
		20	175/0.0641/9582/4967	568/0.2525/34044/17591	Fail	Fail
		50	494/0.2568/28350/14670	Fail	Fail	Fail
		60	810/0.6307/47700/24661	Fail	Fail	Fail
Dixon3dq	(-1,...,-1)	2	3/0.0183/52/30	4/0.0253/78/44	3/0.0327/52/30	4/0.0320/78/44
		4	6/0.0231/144/79	8/0.0346/194/106	46/0.0480/1298/696	46/0.0390/1298/696
		10	14/0.0135/340/185	15/0.0158/348/190	323/0.0596/8734/4691	323/0.0560/9336/4992
		100	109/0.0338/2890/1555	114/0.1766/3014/1622	Fail	Fail
Engval 1	(2,...,2)	100	26/0.0395/1026/540	25/0.0759/992/522	68/0.1067/2812/1475	66/0.0520/2658/1396
		1000	32/0.0576/1492/779	36/0.0630/1754/914	70/0.1065/3374/1758	67/0.1010/3132/1634
		2000	37/0.1156/1864/970	41/0.1455/2090/1087	64/0.2367/3018/1574	72/0.2650/3518/1832
		5000	319/9.2798/19408/10024	70/0.9094/3910/2026	88/1.1751/4660/2419	142/2.0400/7944/4115
Explin 1	(0,...,0)	50	133/0.0817/7482/3875	518/0.4505/30778/15908	687/0.1984/39276/20326	462/0.1598/28034/14480
		100	110/0.0690/6060/3141	372/0.3846/22112/11429	1135/6.0656/68310/35291	450/0.2165/27296/14099
		200	104/0.1037/5842/3026	115/0.2517/6560/3396	Fail	357/0.3260/21572/11144
		500	112/0.3141/6382/3304	193/0.7528/11444/5916	Fail	254/0.5716/15270/7890
Extended DENCHNF	(2,0,...,2,0)	100	47/0.0400/2176/1143	54/0.0576/2250/1180	18/0.0470/674/356	18/0.0470/674/356
		1500	49/0.0866/2348/1233	58/0.1190/2414/1266	19/0.0382/712/376	19/0.0290/712/376
		5000	54/0.4257/2394/1257	59/0.4679/2460/1290	20/0.1242/752/397	20/0.0980/752/397
		10000	54/0.6390/2392/1256	59/0.8869/2460/1290	20/0.2328/752/397	20/0.1970/752/397
Extended Rosenbrock	(0.5,-2, ...,0,-2)	1000	79/0.1047/3612/1886	196/0.4303/8432/4413	39/0.0600/1470/775	95/0.1070/4298/2245
		3000	79/0.2702/3616/1888	91/0.3409/4074/2129	39/0.0760/1468/774	95/0.2936/4298/2245
		4000	83/0.4467/3796/1982	195/1.6014/8318/4355	39/0.0918/1462/771	95/0.3950/4298/2245
		5000	83/0.6450/3778/1973	109/0.9498/4792/2506	39/0.1780/1474/777	95/0.5688/4298/2245
Extended BD1	(0.1,...,0.1)	1000	136/0.2398/4748/2511	137/0.2865/4738/2507	26/0.0520/768/411	26/0.0230/768/411
		2000	137/0.4649/4776/2526	137/0.4576/4750/2513	26/0.0412/768/411	26/0.0270/768/411
		3000	137/0.7545/4776/2526	137/0.7490/4754/2515	26/0.0540/768/411	26/0.0400/768/411
		5000	139/1.5880/4948/2617	143/1.6201/5006/2647	27/0.1340/798/427	27/0.1180/798/427
Extended DENCHNB	(0.1,...,0.1)	1000	11/0.0266/274/149	11/0.0286/274/149	17/0.0351/448/242	17/0.0340/448/242
		2000	11/0.0156/274/149	11/0.0159/274/149	17/0.0226/448/242	17/0.0160/448/242
		3000	12/0.0129/304/165	12/0.0156/304/165	17/0.0184/448/242	17/0.0110/448/242
		5000	12/0.0319/304/165	12/0.0356/304/165	17/0.0636/448/242	17/0.0370/448/242

Table 5: Numerical results of ICD, CD, GICY and IFR.

Test problem	Initial point	Dim	ICD	CD	GICY	IFR
			Iter/CPU/NF/NG	Iter/CPU/NF/NG	Iter/CPU/NF/NG	Iter/CPU/NF/NG
Extended Hiebert	(0,...,0)	10	7/0.0349/296/156	7/0.0375/296/156	10/0.0336/482/252	8/0.0367/358/188
		500	7/0.0132/296/156	7/0.0163/296/156	10/0.0145/482/252	8/0.0124/358/188
		1000	7/0.0040/296/156	7/0.0042/296/156	10/0.0072/482/252	8/0.0046/358/188
		5000	7/0.0293/296/156	7/0.0263/296/156	10/0.0387/482/252	8/0.0295/358/188
Extended Himmelblau	(1,...,1)	10	7/0.0251/452/241	14/0.0383/472/251	17/0.0341/570/303	17/0.0250/570/303
		100	7/0.0090/452/241	14/0.0158/472/251	18/0.0130/606/322	18/0.0180/606/322
		1000	8/0.0047/490/261	15/0.0097/510/271	19/0.0074/640/340	19/0.0100/640/340
		10000	10/0.0928/528/281	16/0.1085/546/290	20/0.1582/680/361	20/0.1170/680/361
Extended tridiagonal 1	(2,...,2)	6	134/0.0338/3480/1885	143/0.0451/3542/1915	39/0.0495/934/507	144/0.0587/3504/1897
		100	295/0.0863/7210/3901	271/0.1751/6724/3634	51/0.0275/1212/658	272/0.1000/6650/3598
		1000	497/2.1144/12346/6671	464/2.8146/11540/6235	273/1.0318/6664/3606	468/2.2788/11478/6208
		5000	710/22.0663/17484/945	635/21.7859/15832/8552	450/20.6391/10854/5878	710/44.0586/17394/9408
Extended White and Holst	(-1.2,1, ...,-1.2,1)	1000	8/0.1750/378/198	120/2.2484/5052/2647	15/0.4560/566/299	120/3.9360/5046/2653
		3000	8/0.3882/378/198	121/6.2276/5126/2685	15/2.0550/566/299	121/18.6620/5110/2677
		5000	8/0.6971/378/198	121/10.9134/5132/2688	15/5.0010/566/299	121/45.6460/5110/2677
		10000	8/1.7490/378/198	129/23.0342/5456/2858	15/16.5080/566/299	129/15.7280/5472/2866
Fletchr	(0,...,0)	2	4/0.0167/138/74	4/0.0173/138/74	4/0.0183/138/74	4/0.0330/138/74
		4	47/0.0212/2140/1118	47/0.0317/2112/1104	70/0.0335/3110/1626	70/0.0340/3110/1626
		6	374/0.0548/19094/9922	371/0.0673/19030/9887	126/0.0229/5590/2922	126/0.0180/5652/2953
		10	1210/0.1857/64988/33705	1198/0.1988/64468/33433	361/0.0595/17132/8928	361/0.0330/16026/8375
Generalized Quartic	(1,...,1)	4	12/0.0366/320/173	12/0.0234/312/169	12/0.0294/288/157	12/0.0200/288/157
		10	10/0.0128/240/131	10/0.0189/230/126	10/0.0101/256/139	10/0.0140/248/135
		50	8/0.0010/258/140	11/0.0022/284/154	11/0.0014/276/150	11/0.0020/276/150
		1000	11/0.0048/280/152	11/0.0046/278/151	12/0.0057/304/165	12/0.0070/304/165
Generalized Rosenbrok	(-1.2,1, ...,-1.2,1)	2	74/0.0365/3420/1785	185/0.0535/7974/4173	36/0.0378/1348/711	84/0.0318/3852/2011
		4	76/0.0208/3270/1712	76/0.0234/3212/1683	Fail	75/0.0115/3206/1679
		6	653/0.0911/32394/16851	650/0.1001/32116/16709	Fail	657/0.0432/32710/17013
		10	Fail	Fail	Fail	Fail
Hager	(1,...,1)	10	10/0.0239/322/174	12/0.0366/326/176	21/0.0347/630/337	12/0.0302/322/174
		20	16/0.0113/522/278	16/0.0155/510/272	30/0.0194/1032/547	16/0.0101/518/276
		50	18/0.0047/688/366	21/0.0078/736/390	49/0.0202/1910/1005	21/0.0039/706/375
		150	28/0.0158/1226/644	36/0.0215/1556/815	Fail	30/0.0136/1256/659
Extended penalty QP1	(1,...,1)	50	7/0.0304/344/183	9/0.0360/354/187	19/0.0557/772/406	18/0.0480/758/398
		500	19/0.1069/918/479	27/0.1713/1440/748	48/0.3431/2796/1447	11/0.0540/446/235
		1000	33/0.3001/1750/940	48/0.4405/2760/1429	244/2.7559/14950/7720	85/0.6840/5092/2632
		1500	36/0.3842/1866/967	332/5.2675/20436/10551	23/0.2556/1250/649	23/0.2240/1250/649
extended penalty QP2	(1,...,1)	40	41/0.0489/1912/1000	44/0.0636/1924/1007	Fail	Fail
		60	173/0.2401/8038/4193	184/0.2702/8564/4467	Fail	Fail
		100	169/0.3773/7848/4098	489/1.1408/23676/12328	Fail	Fail
		1000	251/1.006/9444/6567	Fail	Fail	Fail
perturbed quadratic diagonal	(0.5,...,0.5)	2	9/0.0271/206/113	8/0.0400/180/99	11/0.0160/248/136	39/0.0340/880/480
		50	38/0.0307/1144/613	42/0.0386/1186/636	180/0.1250/5764/3063	Fail
		100	64/0.0530/2016/1073	60/0.0778/1822/972	327/0.484/11272/5964	Fail
		1000	213/1.7090/8038/4233	194/1.5051/7248/3819	Fail	Fail
perturbed quadratic	(0.5,...,0.5)	2	5/0.0177/120/66	5/0.0142/120/66	12/0.0278/320/173	12/0.0270/320/173
		50	41/0.0407/1456/772	44/0.0497/1490/790	350/0.4131/13106/6904	350/0.2380/13072/6887
		100	65/0.0741/2300/1214	68/0.1580/2478/1308	696/1.3605/25384/13389	698/0.9960/27648/14523
		1000	217/4.2883/8778/4597	207/4.6026/8764/4590	Fail	Fail
Raydan 1	(1,...,1)	2	6/0.0312/80/47	6/0.0241/78/46	10/0.0279/176/99	10/0.0320/176/99
		6	10/0.0129/344/188	15/0.0139/382/207	36/0.0144/888/481	36/0.0190/896/485
		70	58/0.0240/2068/1093	57/0.0208/2038/1077	375/0.1286/14322/7537	390/0.0590/15076/7929
		100	63/0.0256/2486/1313	69/0.0286/2638/1389	515/0.1629/19362/10197	526/0.0870/20074/10564
Sine	(1,...,1)	2	3/0.0128/76/42	3/0.0135/76/42	3/0.0164/76/42	3/0.0160/76/42
		4	12/0.0178/624/325	180/0.0992/11040/5701	Fail	42/0.0240/2160/1123
		6	15/0.0124/810/421	134/0.0757/8176/4223	Fail	192/0.0390/11550/5968
		10	14/0.0039/736/383	120/0.0630/7312/3777	Fail	194/0.0160/11738/6064
Staircase2	(0,...,0)	10	22/0.0341/556/300	21/0.0355/554/299	259/0.0686/7548/4034	258/0.0640/7442/3980
		100	191/0.0250/4992/2685	188/0.0865/4964/2671	Fail	Fail
		500	783/1.4704/21034/11301	956/4.0893/26018/13966	Fail	Fail
		1000	1195/7.1915/32738/17565	1981/25.1112/52716/28340	Fail	Fail
Extended Penalty	(1,...,1)	10	7/0.0222/344/184	11/0.0345/382/203	14/0.0437/496/263	14/0.0350/502/266
		100	8/0.0259/350/185	9/0.0268/356/188	26/0.0723/1110/582	25/0.0480/982/517
		500	10/0.0625/424/223	7/0.0314/264/140	23/0.1653/1228/638	19/0.1020/984/512
		1000	9/0.0582/430/225	22/0.1914/1192/619	50/0.4459/2874/1488	19/0.1430/950/495
Quartc	(2,...,2)	2	32/0.0238/676/371	68/0.0209/1468/803	70/0.0269/1512/827	36/0.0320/764/419
		50	59/0.0343/1270/695	195/0.0735/4262/2327	198/0.0812/4328/2363	65/0.0280/1402/767
		500	63/0.1149/1358/743	419/0.8905/9190/5015	422/0.8931/9256/5051	99/0.1720/2150/1175
		5000	95/1.8042/2062/1127	900/18.7540/19772/10787	904/18.8856/19860/10835	148/2.8520/3228/1763

Table 6: Numerical results of ICD, CD, GICY and IFR.

Test problem	Initial point	Dim	ICD	CD	GICY	IFR
			Iter/CPU/NF/NG	Iter/CPU/NF/NG	Iter/CPU/NF/NG	Iter/CPU/NF/NG
Quadratic QF2	(0.5,...,0.5)	10	50/0.0343/1918/1010	50/0.0494/1892/997	87/0.0399/3198/1687	87/0.0340/3228/1702
		100	163/0.0585/7474/3904	168/0.0666/7502/3920	788/0.4563/35972/18775	786/0.1750/34782/18178
		500	381/0.4230/18798/9781	434/0.8196/21966/11418	Fail	Fail
		5000	1225/55.5933/66456/34454	Fail	Fail	Fail
Extended Maratos	(1.1,0.1, ...,1.1,0.1)	100	1634/1.3725/99552/51411	1264/1.4176/75476/39003	41/0.0309/1742/913	Fail
		700	1634/1.3725/99552/51411	1243/9.6540/73804/38146	166/0.2452/9524/4929	Fail
		1000	1425/11.673/86188/44520	1599/20.4783/94648/48924	137/0.2398/7688/3982	Fail
		1500	1361/16.0721/81360/42042	1232/18.2823/74134/38300	Fail	Fail
Extended Freudenstein and Roth	(0.5,-2, ...,0.5,-2)	100	4/0.0234/146/78	4/0.0344/146/98	4/0.0472/146/78	4/0.0350/146/78
		5000	4/0.2192/146/78	4/0.2247/146/78	4/0.2464/146/78	4/0.2030/146/78
		10000	4/0.4053/146/78	4/0.4078/146/78	4/0.4156/146/78	4/0.4050/146/78
		50000	4/3.3621/146/78	4/1.8584/146/78	4/1.9449/146/78	4/1.8670/146/78
Extended quadratic exponential EP1	(1.5,...,1.5, ...,1.1,0.1)	1000	74/0.2210/4508/2329	5/0.0308/214/113	62/0.2196/3746/1936	4/0.0280/150/80
		5000	23/0.2877/1346/697	5/0.0664/212/112	113/1.8656/6884/3556	58/0.7380/3458/1788
		10000	29/0.6997/1708/884	805/62.8681/49762/25687	202/7.3323/12428/6417	59/1.3870/3520/1820
		20000	73/3.2735/4428/2288	18/0.6351/998/518	93/4.4221/5672/2930	73/2.9200/4388/2268
BIGGSB1	(0,...,0)	2	4/0.0202/78/44	3/0.0263/52/30	3/0.0255/52/30	4/0.0160/78/44
		4	19/0.0246/1028/534	13/0.0242/708/368	29/0.0326/1692/876	30/0.0370/1756/909
		6	42/0.0512/5938/2333	38/0.0217/2216/1147	30/0.0139/1798/930	40/0.0230/2418/1250
		10	111/3.0786/10938/5432	Fail	Fail	Fail
Power	(1,...,1)	2	8/0.0345/434/226	8/0.0415/356/187	14/0.0400/806/418	14/0.0430/806/418
		50	79/0.0395/4836/2498	110/0.0580/6758/3490	335/0.1383/20708/10690	94/0.0270/5766/2978
		100	118/0.0412/7254/3746	189/0.0941/11656/6018	670/0.5284/41478/21410	140/0.0270/8618/4450
		700	535/3.3437/33108/17090	1334/15.8425/82646/42658	Fail	354/1.0170/21886/11298
Extended PSC1	(3,0.1, ...,3,0.1)	10	16/0.0453/486/260	16/0.0458/480/257	15/0.0390/458/245	15/0.0430/426/229
		100	16/0.0242/522/278	17/0.0355/550/293	16/0.0211/486/260	16/0.0360/520/277
		1000	21/0.0477/800/422	24/0.0598/1014/532	17/0.0329/646/341	17/0.0420/660/348
		10000	49/6.5515/2686/1393	92/13.6157/5306/2746	468/122.5422/28646/14792	94/13.0080/5458/2824
Staircase 1	(1,...,1)	4	10/0.0474/334/181	14/0.0525/378/204	13/0.0224/338/183	68/0.0470/1960/1049
		10	25/0.0140/820/438	30/0.0248/920/491	917/0.1956/33024/17430	916/0.1100/33328/17581
		50	126/0.0293/4878/2566	138/0.0611/5458/2868	124/0.0130/4138/2194	Fail
		100	246/0.1114/10500/5497	292/0.2671/12826/6706	Fail	Fail
HIMMELH	(1.5,...,1.5)	50	5/0.0161/132/74	7/0.0342/172/94	6/0.0342/110/62	6/0.0290/110/62
		500	5/0.0200/132/74	7/0.0204/172/94	6/0.0141/110/62	6/0.0150/110/62
		10000	6/0.1167/154/86	7/0.1224/172/94	6/0.0720/110/62	6/0.0640/110/62
		50000	6/0.4043/154/86	7/0.4673/172/94	6/0.2595/110/62	6/0.2480/110/62
LIARWHD	(4,...,4)	5000	16/0.5540/694/364	35/1.5517/1682/877	23/7.3730/1008/528	36/1.3645/1772/923
		10000	35/1.5128/1706/889	44/13.7127/2302/1196	27/32.1780/1252/654	88/120.7230/4708/2443
		20000	91/9.5620/4902/2543	50/20.7849/2646/1374	32/15.2322/1552/809	44/22.8001/2332/1211
		30000	41/17.2074/2208/1146	Fail	33/38.9186/1642/855	Fail
NONSCOMP	(3,...,3)	100	32/0.0511/1076/571	38/0.0629/1346/712	113/0.1263/4026/2127	113/0.0630/4304/2266
		500	34/0.2677/1164/617	35/0.2660/1240/656	115/1.3502/4186/2209	115/0.8720/4126/2179
		10000	35/0.3898/1224/648	32/0.3890/1100/583	117/2.2397/4244/2240	117/1.8030/4250/2243
		50000	38/1.5324/1340/709	34/1.6110/1210/640	130/8.7047/4236/2249	130/7.7110/4256/2259
DQDRTIC	(3,...,3)	100	14/0.0369/536/285	16/0.0411/552/293	1508/2.4122/60974/31996	1496/1.2510/60554/31774
		500	12/0.0310/494/262	13/0.0161/446/237	1354/11.9161/53734/28222	1336/7.2950/53110/27892
		1000	15/0.0113/544/288	14/0.0157/498/264	1414/16.5482/56880/29855	1396/11.2680/55554/29174
		50000	11/0.5086/446/237	13/0.5098/472/250	Fail	Fail
Raydan 2	(1,...,1)	50	4/0.0227/112/61	4/0.0267/112/61	4/0.0281/112/61	4/0.0250/112/61
		500	4/0.0260/112/61	4/0.0194/112/61	4/0.0228/112/61	4/0.0160/112/61
		5000	5/0.0882/174/93	5/0.0914/174/93	5/0.0892/174/93	5/0.0750/174/93
		10000	6/0.1935/236/125	5/0.1443/174/93	5/0.1523/174/93	5/0.1180/174/93
Quadratic QF1	(1,...,1)	10	17/0.0546/514/275	17/0.0545/456/246	72/0.0920/2196/1171	72/0.0580/2292/1219
		50	37/0.0465/1440/765	44/0.0558/1494/792	356/0.3717/11592/6153	354/0.2460/12350/6530
		100	61/0.0757/2228/1178	68/0.1864/2362/1250	706/1.5670/23534/12474	706/0.9790/25488/13451
		500	146/0.5699/5722/3008	155/0.9863/6018/3165	Fail	Fail
ARWHEAD	(1,...,1)	50	3/0.0176/116/63	4/0.0302/124/67	4/0.0256/120/65	4/0.0180/120/65
		500	4/0.0191/134/73	5/0.0264/186/99	19/0.0346/1050/545	11/0.0250/554/289
		2000	6/0.0091/186/99	34/0.1052/1984/1027	20/0.0474/1112/577	15/0.0400/802/417
		50000	91/5.4656/5510/2847	28/2.7753/1588/823	711/265.3126/43954/22689	128/16.3430/7790/4024
Nondia	(-1,...,-1)	2	3/0.0161/114/62	4/0.0151/140/75	4/0.0235/140/75	4/0.0160/114/62
		1000	6/0.0208/270/142	17/0.0265/862/449	18/0.0327/884/461	16/0.0297/812/423
		10000	0.0694/392/205	11/0.1479/588/306	225/8.5570/12474/6463	16/0.2390/854/444
		50000	19/1.1500/1230/638	24/1.2723/1368/709	Fail	23/2.4789/1300/674

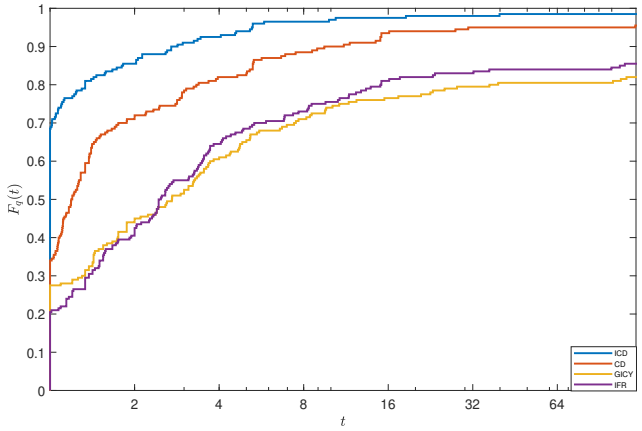


Figure 5: Performance profiles plot based on the number of iterations.

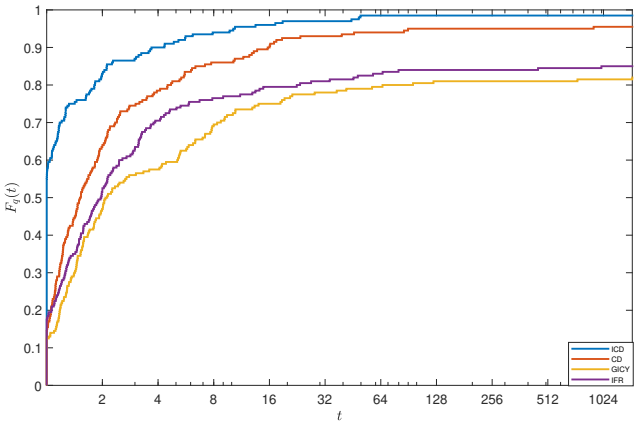


Figure 6: Performance profiles plot based on CPU time.

Figures 5–8 illustrate that the ICD method outperforms the others; notably, it is faster for approximately 60% of the test problems and successfully solves around 98% of them, followed by the CD method with 95% of test problems. IFR method has solved around 85% of the test problems, whereas the GICY method has solved about 82% of the test problems. These outcomes prove the competitiveness and rapid convergence of the ICD algorithm in the majority of the test problems.

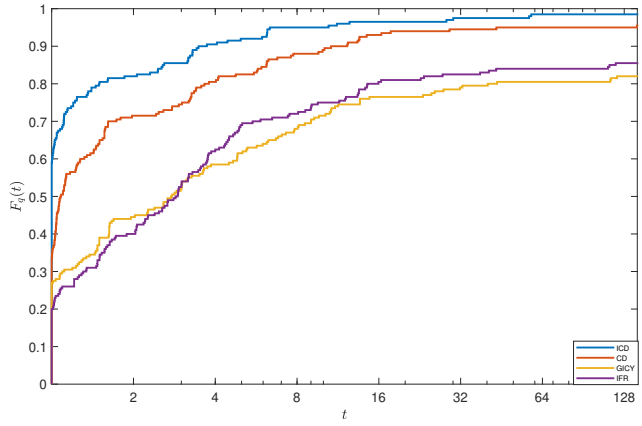


Figure 7: Performance profiles plot based on the number of function evaluations.

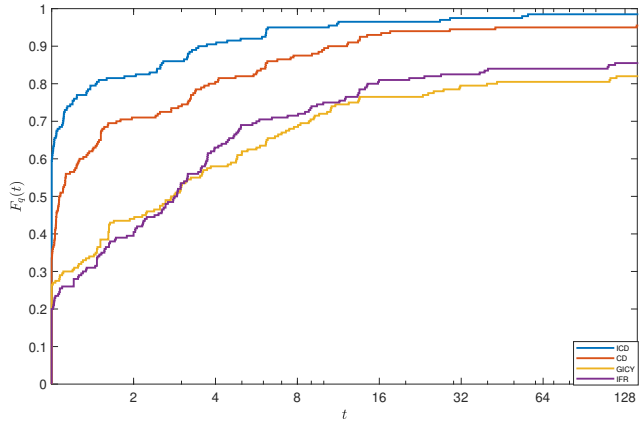


Figure 8: Performance profiles plot based on the number of gradient evaluations.

6 Conclusion

This paper introduced an improvement of the CD method, named ICD, for training two-layer neural network models as well as for solving large-scale nonlinear unconstrained minimization problems. The convergence properties of the proposed ICD approach were analyzed under the strong Wolfe conditions, and global convergence is guaranteed under mild assumptions. To validate the practical efficiency of the proposed method, we compared its performance with three similar variants: CD, GICY, and IFR. Numerical comparisons were conducted across 10 neural network models (ranging from 32 to 4080 weights) and 200 test functions. The results

indicate that the proposed method is highly effective, outperforming these approaches on the majority of test problems.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Andrei, N., *An unconstrained optimization test functions collection*, Adv. Model. Optim. 10 (2008), 147–161.

[2] Belciug, S., *Logistic regression paradigm for training a single-hidden layer feedforward neural network. Application to gene expression datasets for cancer research*, J. Biomed. Inform. 102 (2020), 103373.

[3] Dai, Y.H., Han, J.Y., Liu, G.H., Sun, D.F., Yin, X., and Yuan, Y., *Convergence properties of nonlinear conjugate gradient methods*, SIAM J. Optim. 10 (1999), 348–358.

[4] Dai, Y.H., and Yuan, Y., *A nonlinear conjugate gradient method with strong global convergence properties*, SIAM J. Optim. 10 (1999), 177–182.

[5] Dolan, E.D., and Moré, J.J., *Benchmarking optimization software with performance profiles*, Math. Program. 91 (2002), 201–213.

[6] Fletcher, R., and Reeves, C.M., *Function minimization by conjugate gradients*, Comput. J. 7 (1964), 149–154.

[7] Fletcher, R., *Practical Methods of Optimization*, 2nd ed., John Wiley & Sons, New York, 1987.

[8] Gilbert, J.C., and Nocedal, J., *Global convergence properties of conjugate gradient methods for optimization*, SIAM J. Optim. 2 (1992), 21–42.

[9] Guefassa, I., and Chaib, Y., *A modified conjugate gradient method for unconstrained optimization with application in regression function*, RAIRO-Oper. Res. 59 (2025), 311–324.

[10] Hestenes, M.R., and Stiefel, E.L., *Methods of conjugate gradients for solving linear systems*, J. Res. Natl. Bur. Stand. 49 (1952), 409–436.

[11] Huang, G.B., Zhu, Q.Y., and Siew, C.K., *Extreme learning machine: theory and applications*, Neurocomputing, 70 (2006), 489–501.

[12] Jiang, X.Z., and Jian, J.B., *Improved Fletcher-Reeves and Dai-Yuan conjugate gradient methods with the strong Wolfe line search*, J. Comput. Appl. Math. 348 (2019), 525–534.

[13] Jiang, X., Pan, L., Liu, M., and Jian, J., *A family of spectral conjugate gradient methods with strong convergence and its applications in image restoration and machine learning*, J. Franklin Inst. 361 (2024), 107033.

[14] Lagaris, I.E., Likas, A., and Fotiadis, D.I., *Artificial neural networks for solving ordinary and partial differential equations*, IEEE Trans. Neural Netw. 9 (1998), 987–1000.

[15] Liu, J.K., and Li, S.J., *New hybrid conjugate gradient method for unconstrained optimization*, Appl. Math. Comput. 245 (2014), 36–43.

[16] Liu, Y., and Storey, C., *Efficient generalized conjugate gradient algorithms, part 1: theory*, J. Comput. Theory Appl. 69 (1991), 129–137.

[17] Maiza, A., Ziadi, R., Saleh, M.A., and Almaymuni, A.Z., *An improved conjugate gradient algorithm by adapting a new line search technique*, Electron. Res. Arch. 33 (2025), 2148–2171.

[18] McDonnell, M.D., Tissera, M.D., Vladusich, T., van Schaik, A., and Tapson, J., *Fast, simple and accurate handwritten digit classification by training shallow neural network classifiers with the extreme learning machine algorithm*, PLoS ONE 10 (2015), e0134254.

[19] Polak, E., and Ribière, G., *Note sur la convergence de méthodes de directions conjuguées*, Rev. Française Informat. Recherche Opérationnelle, 3(R1) (1969), 35–43.

[20] Polyak, B.T., *The conjugate gradient method in extreme problems*, USSR Comput. Math. Math. Phys. 9 (1969), 94–112.

[21] Rivaie, M., Mamat, M., June, L.W., and Mohd, I., *A new class of nonlinear conjugate gradient coefficients with global convergence properties*, Appl. Math. Comput. 218 (2012), 11323–11332.

[22] Souli, C., Ziadi, R., Bencherif-Madani, A., and Khudhur, H.M., *A hybrid CG algorithm for nonlinear unconstrained optimization with application in image restoration*, J. Math. Model. 12 (2024), 301–317.

[23] Souli, C., Ziadi, R., Lakhdari, I., and Leulmi, A., *An efficient hybrid conjugate gradient method for unconstrained optimization and image restoration problems*, Iran. J. Numer. Anal. Optim. 15 (2025), 99–123.

[24] Touati-Ahmed, D., and Storey, C., *Efficient hybrid conjugate gradient technique*, J. Optim. Theory Appl. 64 (1990), 379–397.

[25] Wei, Z., Yao, S., and Liu, L., *The convergence properties of some new conjugate gradient methods*, Appl. Math. Comput. 183 (2006), 1341–1350.

[26] Yousif, O.O.O., Ziadi, R., Almaymuni, A.Z., and Saleh, M.A., *An improved version of Polak-Ribière-Polyak conjugate gradient method with its applications in neural networks training*, Electron. Res. Arch. 33 (2025), 4799–4815.

[27] Zhang, L., *An improved Wei-Yao-Liu nonlinear conjugate gradient method for optimization computation*, Appl. Math. Comput. 215 (2009), 2269–2274.

[28] Zhu, Z., Zhang, D., and Wang, S., *Two modified DY conjugate gradient methods for unconstrained optimization problems*, Appl. Math. Comput. 373 (2020), 125004.

[29] Ziadi, R., Maiza, A., Mehamdia, A.E., and Saleh, M.A., *An enhanced PRP conjugate gradient method by adapting a new line search for unconstrained optimization and image restoration problems*, Appl. Set-Valued Anal. Optim. 8 (2026), 357–372.

Iranian Journal of Numerical Analysis and Optimization
Vol. 17, No. 1, 2027, pp 359–379
https://doi.org/10.22067/ijnao.2026.97071.1808
https://ijnao.um.ac.ir/

CC BY NC

Research Article

Enhancing conjugate gradient performance via a diagonal quasi-Newton strategy: Applications in adaptive filtering and power allocation

N. Bidabadi*,

Abstract

Unconstrained optimization is fundamental to scientific computing. While nonlinear conjugate gradient (CG) methods are valued for their low memory requirements and simplicity, their convergence may be slow. Conversely, quasi-Newton methods like Broyden–Fletcher–Goldfarb–Shanno (BFGS) achieve faster superlinear convergence but become computationally prohibitive for large-scale problems. To address this trade-off, this paper introduces a novel hybrid algorithm that combines a diagonal quasi-Newton approach with a CG method. This framework leverages the complementary strengths of both techniques, maintaining modest memory usage while accelerating convergence. Numerical experiments on several benchmark problems, adaptive filter design and power allocation, demonstrate the proposed method's effectiveness. Our MATLAB-based evaluations show that the hybrid approach consistently outperforms standard CG and BFGS methods.

AMS subject classifications (2020): Primary 90C30; Secondary 65K05.

Keywords: Conjugate gradient; Diagonal quasi-Newton; Adaptive filtering; Power allocation

*Corresponding author

Received 18 December 2025; revised 18 February 2026; accepted 27 March 2026

Narges Bidabadi
Department of Mathematical Sciences, Yazd University, Yazd, Iran. e-mail: n_bidabadi@yazd.ac.ir

How to cite this article
Bidabadi, N., Enhancing conjugate gradient performance via a diagonal quasi-Newton strategy: Applications in adaptive filtering and power allocation. *Iran. J. Numer. Anal. Optim.*, 2027; 17(1): 359-379. https://doi.org/10.22067/ijnao.2026.97071.1808

1 Introduction

Unconstrained optimization, the problem of minimizing a nonlinear function

$$\min f(x), \quad x \in \mathbb{R}^n, \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable, forms the computational backbone of countless applications in engineering, data science, and computational physics. As the scale of these problems continues to increase, the development of efficient, scalable, and robust iterative algorithms remains a central pursuit in numerical optimization. Iterative methods for minimization are broadly characterized by their trade-off between computational cost per iteration and convergence rate. Two of the most prominent families of algorithms in this domain are quasi-Newton methods and nonlinear conjugate gradient (CG) methods.

Quasi-Newton methods emerged in the 1950s and 1960s as a revolutionary class of algorithms for unconstrained optimization, bridging the gap between simple gradient methods and computationally expensive Newton's method. The fundamental idea is to approximate the Hessian matrix (or its inverse) using only gradient information, thus avoiding the costly computation of second derivatives.

The first significant breakthrough came with the Davidon–Fletcher–Powell method proposed by Davidon [6] in 1959 and later popularized by Fletcher and Powell [10] in 1963. This method introduced the concept of updating an inverse Hessian approximation iteratively through rank-two updates that satisfy the secant condition.

However, the most influential development came with the Broyden–Fletcher–Goldfarb–Shanno (BFGS) method, which was independently developed around 1970 [4, 9, 11, 20]. BFGS quickly became the gold standard for quasi-Newton methods due to its numerical stability and superior performance in practice. The BFGS update maintains positive definiteness of the Hessian approximation under appropriate conditions and demonstrates robust convergence behavior.

A major weakness of the BFGS method is that as the dimension of the problem increases, its computational cost and memory requirements also increase. Furthermore, especially if the approximation of the Hessian matrix becomes ill-conditioned, a suitable solution may not be obtained. Numerous diagonal quasi-Newton (DQN) methods have been proposed in recent years to address these issues. For example, Enshaei, Leong, and Farid [8] (2016) proposed a DQN

method for solving unconstrained optimization problems. Assuming that the Hessian matrix at iteration k is diagonal and positive definite, they derived the Hessian matrix for the next iteration such that it has the shortest distance from the current iteration matrix in the Frobenius norm while satisfying the weak secant condition. Similarly, Sim et al. [21] developed a diagonally updated quasi-Newton formula using the log-determinant norm, ensuring that it satisfies the weak secant equation. More recently, in 2024, an accelerated structured DQN method was proposed for solving nonlinear least squares problems [13], followed in 2025 by a DQN method designed for nonlinear equations [5].

A fundamental contribution to optimization theory was made by Polyak [18], who systematically analyzed the CG method for extremal problems. While nonlinear CG methods are valued for their low memory requirements and simplicity, they are often hampered by slow convergence and sensitivity to problem conditioning. This stems from their reliance on gradient information alone, which can lead to inefficient zigzagging search paths, especially on ill-conditioned or non-quadratic landscapes.

To address this weakness, many improvements to the CG algorithm have been proposed. For example, in 2009, Andrei proposed a fundamentally different CG method in which the conjugacy parameter is computed by approximating the Hessian/vector product through finite differences. Additionally, for the step length computation, he suggested an acceleration scheme capable of improving the efficiency of the algorithm [3]. More recently, Zhu, Zhu, and Tan [22] proposed an adaptive two-parameter accelerated CG method that satisfies the sufficient descent condition for the search direction.

The complementary strengths and weaknesses of these two families suggest a compelling opportunity for a hybrid approach. Aini et al. [1] used the quasi-Newton direction instead of the steepest descent direction when calculating the search direction in the CG method. In 2023, Osinuga and Stanimirovic [17] proposed a modification of hybrid CG methods by appropriately combining the BFGS update with hybridized CG parameters.

In this paper, we propose a new hybrid algorithm that integrates a DQN method and a CG method. The rest of this paper is organized as follows: Section 2 reviews necessary background on CG and quasi-Newton methods. Section 3 details the derivation and mechanics of our proposed hybrid algorithm. Section 4 presents the convergence analysis. Section 5 presents the numerical results, and Section 6 concludes the paper.

2 Preliminaries

Iterative methods for solving problem (1), by using initial point x_0, generate a sequence of iterates x_k according to the update formula:

$$x_{k+1} = x_k + \alpha_k d_k, \tag{2}$$

where $\alpha_k > 0$ denotes the step length and d_k represents the search direction at the kth iteration. In quasi-Newton methods, the search direction d_k is computed by solving the linear system:

$$B_k d_k = -g_k, \tag{3}$$

where $g_k = \nabla f(x_k)$ is the gradient and B_k is a symmetric matrix that approximates the Hessian of f at x_k. The construction of B_k relies on the secant condition, which requires the matrix to satisfy

$$B_k s_{k-1} = y_{k-1}, \tag{4}$$

where $s_{k-1} = x_k - x_{k-1}$ and $y_{k-1} = g_k - g_{k-1}$. This condition ensures that B_k mimics the behavior of the true Hessian along s_{k-1}.

Among the most important quasi-Newton methods is the BFGS method, in which the Hessian matrix is updated at each iteration using the following relation:

$$B_{k+1} = B_k + \frac{y_k y_k^\top}{y_k^\top s_k} - \frac{B_k s_k s_k^\top B_k}{s_k^\top B_k s_k}. \tag{5}$$

In contrast, the nonlinear CG method computes the search direction recursively as

$$d_k = \begin{cases} -g_k & k = 0, \\ -g_k + \beta_{k-1} d_{k-1}, & k = 1, 2, \ldots, \end{cases} \tag{6}$$

where the scalar β_k is a conjugacy parameter that distinguishes different CG variants. The Rivaie–Mamat–Ismail–Leong (RMIL) formula is given by [19]

$$\beta_{k-1} = \frac{g_k^T y_{k-1}}{d_{k-1}^T d_{k-1}}. \tag{7}$$

Note: The step length $\alpha_k \in (0, \alpha_{\max})$ in this study is determined by the Wolfe conditions [16]:

$$\begin{aligned} f(x_k + \alpha_k d_k) &\leq f(x_k) + c_1 \alpha_k g_k^T d_k, \\ \nabla f(x_k + \alpha_k d_k)^T d_k &\geq c_2 \nabla f(x_k)^T d_k, \end{aligned} \tag{8}$$

with $0 < c_1 < c_2 < 1$. Ensuring that the line search satisfies the Wolfe conditions guarantees that $s_k^T y_k > 0$ for all k [16].

3 A combined CG and DQN approach

The search direction d_k in our hybrid algorithm is determined through an adaptive two-case strategy:

Case 1: DQN phase

Under the condition that $s_{k-1}^T y_{k-1} - \|s_{k-1}\|^2 > 0$, the diagonal positive definite matrix $B_k = Diag(b_k^1, \ldots, b_k^n)$ is built such that it satisfies the weak secant equation

$$s_{k-1}^T B_k s_{k-1} = s_{k-1}^T y_{k-1}, \tag{9}$$

while guaranteeing that the diagonal entries of the matrix avoid becoming excessively small, thus maintaining numerical stability. The matrix B_k is determined by solving the optimization problem

$$\begin{aligned} &\min \ trace(B_k) = b_k^1 + \cdots + b_k^n \\ &s.t. \ s_{k-1}^T B_k s_{k-1} = s_{k-1}^T y_{k-1}, \\ &\quad b_k^i \geq 1, \quad i = 1, \ldots, n. \end{aligned} \tag{10}$$

The problem (10) is equivalent to

$$\begin{aligned} &\min \tilde{b_k}^1 + \cdots + \tilde{b_k}^n \\ &s.t. \sum_{i=1}^{n} (s_{k-1}^i)^2 \tilde{b_k}^i = s_{k-1}^T y_{k-1} - \|s_{k-1}\|^2, \end{aligned}$$

$$\tilde{b_k}^i \geq 0, \quad i = 1, \ldots, n, \tag{11}$$

where $\tilde{b_k}^i = b_k^i - 1$. This optimization problem bears structural similarities to the knapsack problem and admits an optimal solution given by

$$\tilde{b_k}^i = \begin{cases} \frac{s_{k-1}^T y_{k-1} - \|s_{k-1}\|^2}{(s_{k-1}^j)^2}, & \text{if } j = \arg\max\{(s_{k-1}^i)^2\}, \\ 0, & o.w. \end{cases} \tag{12}$$

Therefore, the optimal solution to problem (10) is characterized by

$$b_k^i = \begin{cases} \frac{s_{k-1}^T y_{k-1} - \|s_{k-1}\|^2}{(s_{k-1}^j)^2} + 1, & \text{if } j = \arg\max\{(s_{k-1}^i)^2\}, \\ 1, & o.w. \end{cases} \tag{13}$$

Case 2: CG phase

When $s_{k-1}^T y_{k-1} - \|s_{k-1}\|^2 \leq 0$, we use the CG-RMIL method; that is, the search direction is calculated by (6), where β_{k-1} is computed by (7).

Consequently, the new hybrid algorithm is as follows.

Algorithm 1 A combined CG and DQN algorithm (CG-DQN)

Require: a starting point x_0, a stopping parameter $\epsilon > 0$, and constants c_1 and c_2 satisfying $0 < c_1 < c_2 < 1$.

1:
2: Set $k = 0$. Calculate $g_0 = \nabla f(x_0)$. If $\|g_0\| < \epsilon$, stop. Otherwise set $d_0 = -g_0$ and go to step 4.
3: If $s_{k-1}^T y_{k-1} - \|s_{k-1}\|^2 \geq 0$, calculate b_k^i, $i = 1, \ldots, n$, using relation (13), calculate the search direction d_k using $d_k^i = -\frac{g_k^i}{b_k^i}$ and go to step 5. Otherwise, proceed to the next step.
4: Calculate β_{k-1} and the search direction d_k using (7) and (6).
5: Calculate α_k satisfying (8). Set $x_{k+1} = x_k + \alpha_k d_k$, and $g_{k+1} = \nabla f(x_{k+1})$.
6: If $\|g_{k+1}\| < \epsilon$, stop and accept x_{k+1} as an approximate solution to the problem; otherwise calculate the values of $s_{k-1} = x_k - x_{k-1}$ and $y_{k-1} = g_k - g_{k-1}$. Set $x_k = x_{k+1}$, $g_k = g_{k+1}$ and $k = k + 1$, go to step 2.

4 Convergence analysis

We now analyze the convergence properties of the proposed algorithm. First, we recall an important convergence theorem for iterative methods known as the Zoutendijk theorem.

Theorem 1. [24] Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function bounded from below, and twice continuously differentiable on an open set Ω that contains the sublevel set $\{x : f(x) \leq f(x_0)\}$, where x_0 is the algorithm's starting point. Suppose further that $g = \nabla f$ is Lipschitz continuous on Ω. If the search direction d_k is a descent direction and the step length α_k satisfies the Wolfe conditions, then

$$\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty. \tag{14}$$

For the convergence proof of Algorithm 1, we adopt the following standard assumptions.

Assumption (A1): The level set $\mathcal{L}_0 = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$ is bounded. That is, there exists a constant $M > 0$ such that

$$\|x\| \leq M, \quad \text{for all } x \in \mathcal{L}_0. \tag{15}$$

Assumption (A2): In some neighborhood Ω of $\mathcal{L}_0$, the function f is continuously differentiable, and its gradient is Lipschitz continuous. Namely, there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \leq L\|x - y\|, \quad \text{for all } x, y \in \Omega. \tag{16}$$

A direct consequence of assumptions (A1) and (A2) is that

$$\|x - y\| \leq 2M, \quad \text{for all } x, y \in \mathcal{L}_0, \tag{17}$$

and there exists a constant $\gamma > 0$ such that

$$\|g(x)\| \leq \gamma, \quad \text{for all } x \in \mathcal{L}_0. \tag{18}$$

In the following result, we show that the search direction of Algorithm 1 is a descent direction.

Lemma 1. Under assumption (A2), the direction d_k in Algorithm 1 satisfies the sufficient descent condition, meaning there exists a constant $c \in (0, 1)$ such that

$$g_k^T d_k \leq -c\|g_k\|^2. \tag{19}$$

Proof. In case 1, we have

$$g_k^T d_k = -g_k^T B_k^{-1} g_k \leq -\|g_k\|^2 \sigma_{\min}(B_k^{-1}) = -\frac{\|g_k\|^2}{\|B_k\|},$$

where $\sigma_{\min}$ denotes the smallest singular value of the matrix. Since,

$$\|B_k\| = \max_i b_k^i = \frac{s_{k-1}^T y_{k-1} - \|s_{k-1}\|^2}{(s_{k-1}^j)^2} + 1, \quad j = \arg\max\{(s_{k-1}^i)^2\},$$

and

$$\frac{s_{k-1}^T y_{k-1} - \|s_{k-1}\|^2}{(s_{k-1}^j)^2} \leq \frac{s_{k-1}^T y_{k-1}}{(s_{k-1}^j)^2} \leq \frac{\|s_{k-1}\|\|y_{k-1}\|}{(s_{k-1}^j)^2} \leq \frac{L\|s_{k-1}\|^2}{(s_{k-1}^j)^2} \leq nL,$$

we conclude that $\|B_k\| \leq nL + 1$, we obtain

$$g_k^T d_k \leq -\frac{\|g_k\|^2}{nL+1}.$$

In case 2, we have

$$\begin{aligned} g_k^T d_k &= -g_k^T g_k + \beta_{k-1} d_{k-1}^T g_k = -\|g_k\|^2 + \frac{g_k^T y_{k-1}}{d_{k-1}^T d_{k-1}} d_{k-1}^T g_k \\ &\leq -\|g_k\|^2 + \|g_k\|^2 \frac{\|y_{k-1}\|\|d_{k-1}\|}{\|d_{k-1}\|^2} \\ &\leq -\|g_k\|^2 + \|g_k\|^2 \frac{L\alpha_{k-1}\|d_{k-1}\|\|d_{k-1}\|}{\|d_{k-1}\|^2} \\ &\leq -\|g_k\|^2 (1 - L\alpha_{\max}). \end{aligned}$$

By choosing $\alpha_{\max} < \frac{1}{L}$, we have $1 - L\alpha_{\max} \in (0, 1)$.
Therefore

$$g_k^T d_k \leq -c\|g_k\|^2,$$

with $c = \min(1 - L\alpha_{\max}, \frac{1}{nL+1})$. □

The convergence result is presented in the following theorem.

Theorem 2. Under assumptions (A1) and (A2), the sequence $\{g_k\}$ generated by the CG-DQN algorithm satisfies

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{20}$$

Proof. The proof follows by contradiction. Assuming $\liminf_{k\to\infty} \|g_k\| \neq 0$, implies $\|g_k\| \geq \xi > 0$ for all k. We first show that $\|d_k\|$ is bounded.

In case 1, we have $b_k^i \geq 1$, for all i which yields to $\|B_k^{-1}\| \leq 1$. Thus,

$$\|d_k\| = \|B_k^{-1} g_k\| \leq \|g_k\| \leq \gamma.$$

In case 2, we have

$$\begin{aligned}
\|d_k\| &= \| - g_k + \beta_{k-1} d_{k-1}\| \\
&\leq \|g_k\| + \frac{\|g_k\|\|y_{k-1}\|\|d_{k-1}\|}{\|d_{k-1}\|^2} \\
&\leq \|g_k\| + \frac{\|g_k\| L\|s_{k-1}\|\|d_{k-1}\|}{\|d_{k-1}\|^2} \\
&= \|g_k\| + \frac{\|g_k\| L\alpha_{k-1}\|d_{k-1}\|\|d_{k-1}\|}{\|d_{k-1}\|^2} \\
&\leq \|g_k\|(1 + L\alpha_{\max}) \\
&\leq \gamma(1 + L\alpha_{\max}).
\end{aligned}$$

Since $\|d_0\| = \| - g_0\| \leq \gamma$, we conclude that

$$\|d_k\| \leq \Gamma, \qquad k = 0, 1, \ldots,$$

where

$$\Gamma = \gamma(1 + L\alpha_{\max}).$$

This implies

$$\frac{1}{\|d_k\|} \geq \frac{1}{\Gamma} > 0, \qquad k = 0, 1, \ldots,$$

and

$$\sum_{k=0}^{\infty} \frac{1}{\|d_k\|^2} = +\infty. \tag{21}$$

On the other hand, using relation (19), we have

$$(g_k^T d_k)^2 \geq c^2 \|g_k\|^4 \geq c^2 \xi^4,$$

which together with relation (14) implies that

$$\infty > \sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} \geq \sum_{k=0}^{\infty} \frac{c^2 \xi^4}{\|d_k\|^2} = c^2 \xi^4 \sum_{k=0}^{\infty} \frac{1}{\|d_k\|^2},$$

which directly contradicts condition (21). □

5 Numerical results

A numerical comparative analysis was conducted to evaluate the performance of the proposed CG-DQN algorithm against the CG-RMIL and BFGS methods. In the BFGS method, the search direction d_k is calculated by (3), where B_k is updated by (5) and the initial matrix was taken as the identity matrix. Also, we solve system (3) using a modified Cholesky decomposition and therefore the resulting search direction will always be a descent direction. All algorithms were implemented in MATLAB R2020b and executed on a system with a 3.3 GHz Intel Core i3 processor and 16 GB of RAM. For a fair comparison, we used a Wolfe line search with parameters $c_1 = 0.2$ and $c_2 = 0.8$. The maximum number of iterations was set to 200. The numerical results are presented and analyzed in the following three subsections.

5.1 Standard benchmarks

We evaluated the algorithms on 102 test problems (see [2]) across multiple dimensions (n). Each run was stopped upon satisfying any of the following conditions:

- $\|\nabla f(x_k)\| < 10^{-4}$.
- $\|x_{k+1} - x_k\|_\infty < 10^{-6}$.
- $\|x_{k+1} - x_k\|_2 < 10^{-6}$.

When $d_k^T g_k > -10^{-1}\|d_k\|\|g_k\|$, we restarted the search directions of the CG-DQN and CG-RMIL algorithms to the steepest-descent direction. For our comparisons, we used the notion of a performance profile [7] to evaluate and compare the performance of the set of solvers S on a test set P. Assume that we have n_s solvers and n_p problems. We are interested in using the computational time as a performance measure. Suppose $T_{p,s}$ is the computational time required to solve problem p by algorithm s. We compare the performance on problem p by solver s with the best performance by any solver on this problem; that is, we use the performance ratio

$$r_{p,s} := \frac{T_{p,s}}{\min_s T_{p,s}}.$$

We assume that a parameter R_m is chosen so that $R_m \geq r_{p,s}$, for all p and s, and we put $r_{p,s} = R_m$ if and only if algorithm s does not solve problem p. If we define

$$\rho_s(\tau) := \frac{size\{p \in P \ : \ r_{p,s} \leq \tau\}}{n_p},$$

then $\rho_s(\tau)$ is the probability that solver $s \in S$ that a performance ratio $r_{p,s}$ is within a factor $\tau \in \mathbb{R}$ of the best possible ratio. The function ρ_s is the distribution function for the performance ratio. The value of $\rho_s(1)$ is the probability that the solver will win over the rest of the solvers. The Dolan–Moré [7] performance profiles in Figures 1 and 2 compare algorithms based on computational time. They confirm that the proposed method is more efficient than both CG-RMIL and BFGS approaches.

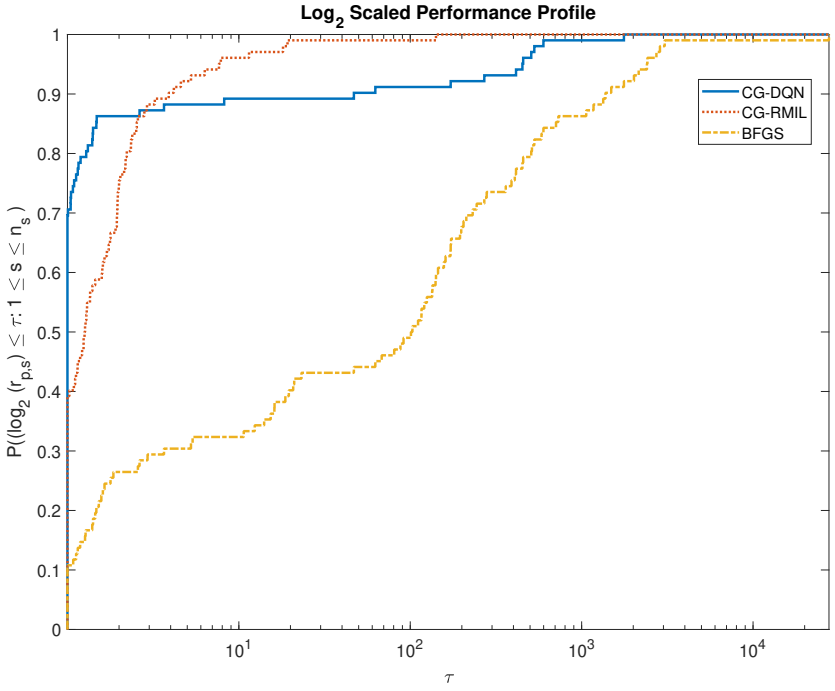


Figure 1: Performance profile based on central processing unit (CPU) time for problems with $n = 500$

5.2 Application to adaptive filter design problem

In this subsection, we apply the proposed algorithm to the fundamental signal processing problem of adaptive filter design. The objective is to perform system identification by finding the optimal coefficients of a finite impulse response (FIR) filter that minimize the prediction error between a desired signal and the filter's output [15, 12].

Problem formulation: The output of an N-tap FIR filter at time t is given by

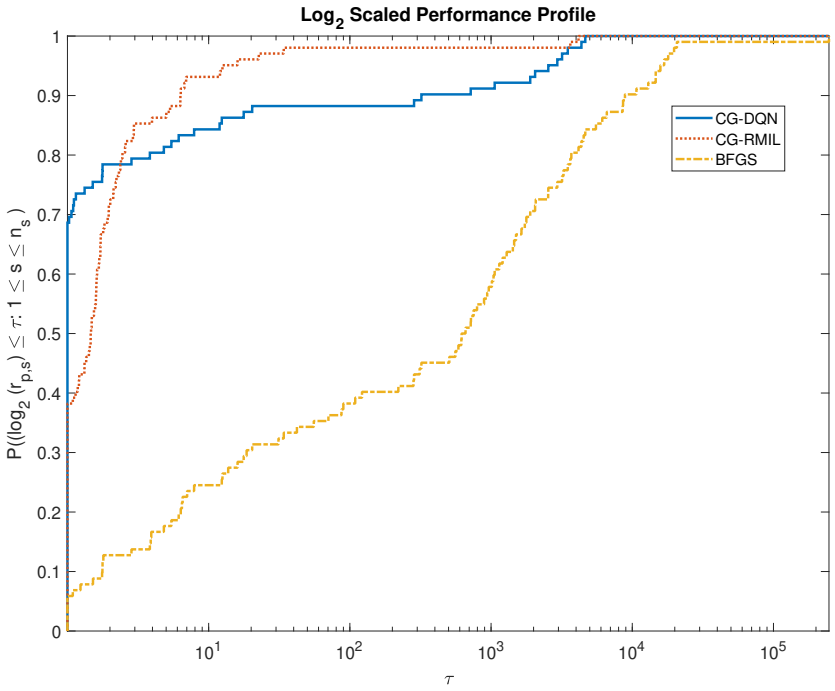


Figure 2: Performance profile based on CPU time for problems with $n = 1000$

$$y[t] = \sum_{k=0}^{N-1} w[k] \cdot x[t-k] = \mathbf{w}^T \mathbf{x}[t],$$

where $\mathbf{w} = [w[0], w[1], \ldots, w[N-1]]^T$ is the vector of filter coefficients (the optimization variables), and $\mathbf{x}[t] = [x[t], x[t-1], \ldots, x[t-N+1]]^T$ is the input vector. The instantaneous error is

$$e[t] = d[t] - y[t],$$

with $d[t]$ being the desired signal. The goal is to minimize the mean square error,

$$\min_{\mathbf{w}} J(\mathbf{w}) = \mathbb{E}\left[|e[t]|^2\right] = \mathbb{E}\left[|d[t] - \mathbf{w}^T \mathbf{x}[t]|^2\right].$$

For practical optimization, we use the sample average approximation over M data points

$$\min_{\mathbf{w} \in \mathbb{R}^N} J(\mathbf{w}) = \frac{1}{M} \sum_{t=N}^{M} \left(d[t] - \mathbf{w}^T \mathbf{x}[t]\right)^2.$$

To simulate a realistic system identification scenario, synthetic data were generated as follows. The input signal $x[t]$ was produced as colored noise using a second-order autoregressive (AR(2)) process:

$$x[t] = 0.5x[t-1] - 0.1x[t-2] + v[t], \quad v[t] \sim \mathcal{N}(0,1), \tag{22}$$

where $\mathcal{N}(0,1)$ is zero-mean, unit-variance Gaussian white noise. The generated signal was then normalized to unit variance:

$$x[t] \leftarrow \frac{x[t]}{\sigma_x}. \tag{23}$$

The desired signal was generated according to the system identification model:

$$d[t] = \mathbf{w}_{\text{Opt}}^T \mathbf{x}[t] + \eta[t], \quad \eta[t] \sim \mathcal{N}(0, \sigma_\eta^2), \tag{24}$$

with noise variance $\sigma_\eta^2 = 0.0004$, corresponding to a signal-to-noise ratio of approximately 34 dB. The true (unknown) system was modeled by a coefficient vector $\mathbf{w}_{\text{Opt}}$ that follows an exponential decay pattern with minor random perturbations:

$$\mathbf{w}_{\text{Opt}} = [0.9, 0.7, 0.5, 0.3, 0.1, \epsilon_6, \epsilon_7, \ldots, \epsilon_N]^T, \tag{25}$$

where $\epsilon_i \sim \mathcal{N}(0, 0.0025)$ for $i = 6, 7, \ldots, N$ represent small random perturbations.
The problem dimensions were set to a filter length of $N = 32$ coefficients, $M = 500000$ data samples, leading to a 32-dimensional optimization over $\mathbf{w} \in \mathbb{R}^{32}$.

We executed the BFGS, CG-RMIL, and CG-DQN algorithms to solve this problem. A common stopping criterion of $\|\nabla J(\mathbf{w})\| < 10^{-4}$ was used for all methods. Figures 3–5 present comparative analyses of the BFGS, CG-RMIL, and proposed CG-DQN algorithms, respectively, through the following subplots:

- Convergence Analysis: shows the exponential decay of both the objective function value and the gradient norm, indicating rapid convergence toward the optimum.

- Filter Coefficients Comparison: compares the estimated filter coefficients against the true optimal values.

- Signal Prediction: demonstrates the filtering performance by plotting the desired signal $d[t]$ alongside the filter output $y[t]$ over a representative sample range. The close alignment and minimal residual error $e[t] = d[t] - y[t]$ confirm the filter's effectiveness in tracking the system behavior.

- Error Distribution: presents a histogram of the prediction errors $e[t]$. The distribution approximates a zero-mean Gaussian profile, indicating successful modeling with unbiased estimation and appropriate model complexity.

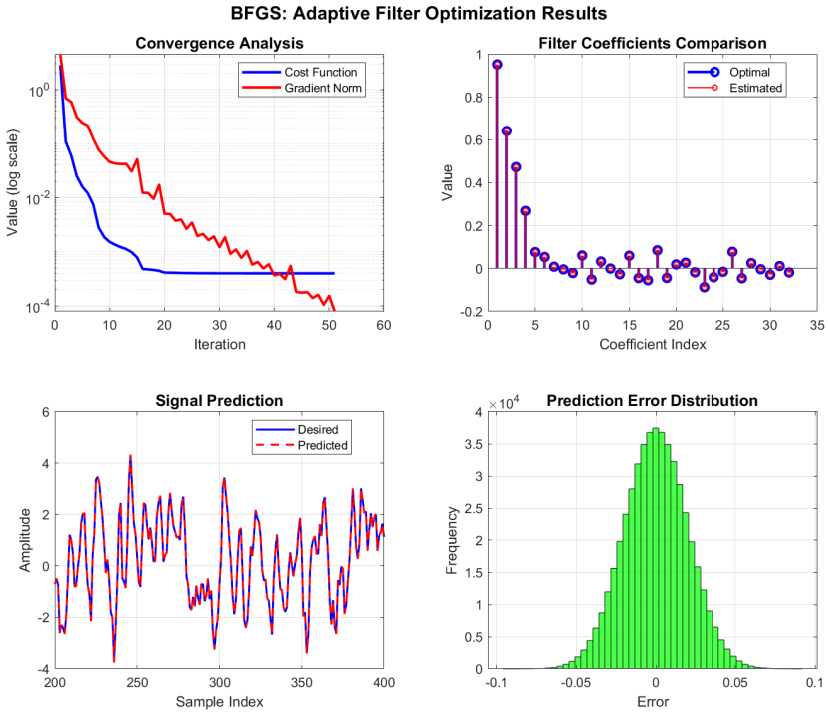


Figure 3: The results obtained from implementing the BFGS algorithm on the adaptive filter design problem.

Collectively, these analyses substantiate the efficacy of the proposed method and show robust convergence, accurate parameter estimation, and excellent predictive performance.

5.3 Application to power allocation problem

This subsection presents the results of applying the proposed algorithm to the power minimization problem in the uplink of a single-input single-output system with K users and one base station. Efficient power allocation is critical in wireless networks to balance energy efficiency with quality-of-service requirements [23, 14].

The problem is formulated as a resource allocation task. Let

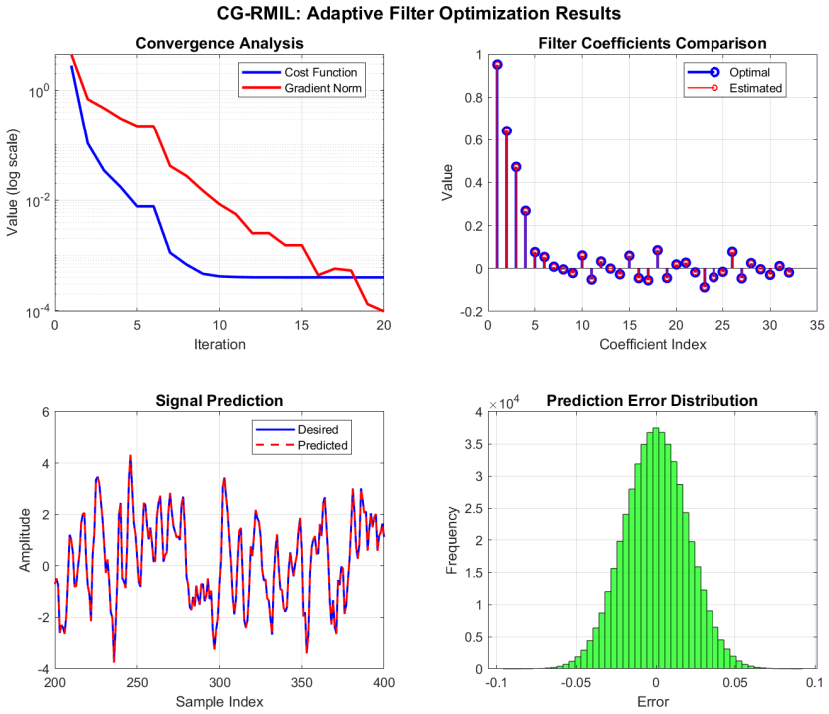


Figure 4: The results obtained from implementing the CG-RMIL algorithm on the adaptive filter design problem.

$$\mathbf{p} = [p_1, p_2, \ldots, p_K], \tag{26}$$

denote the vector of transmit powers for all users. The achievable uplink rate for user i is given by

$$R_i(\mathbf{p}) = \log_2 \left(1 + \frac{p_i |h_{ii}|^2}{\sum_{j \neq i} p_j |h_{ij}|^2 + \sigma^2} \right), \tag{27}$$

where h_{ii} is the direct channel gain between user i and the base station, h_{ij} is the interference channel gain from user j and σ^2 is the noise variance. If $R_{\min}$ is the minimum required rate for user i, and $p_i^{\min}$ and $p_i^{\max}$ denote the minimum and maximum allowable transmit powers for user i, respectively, then the power minimization problem is formulated as follows:

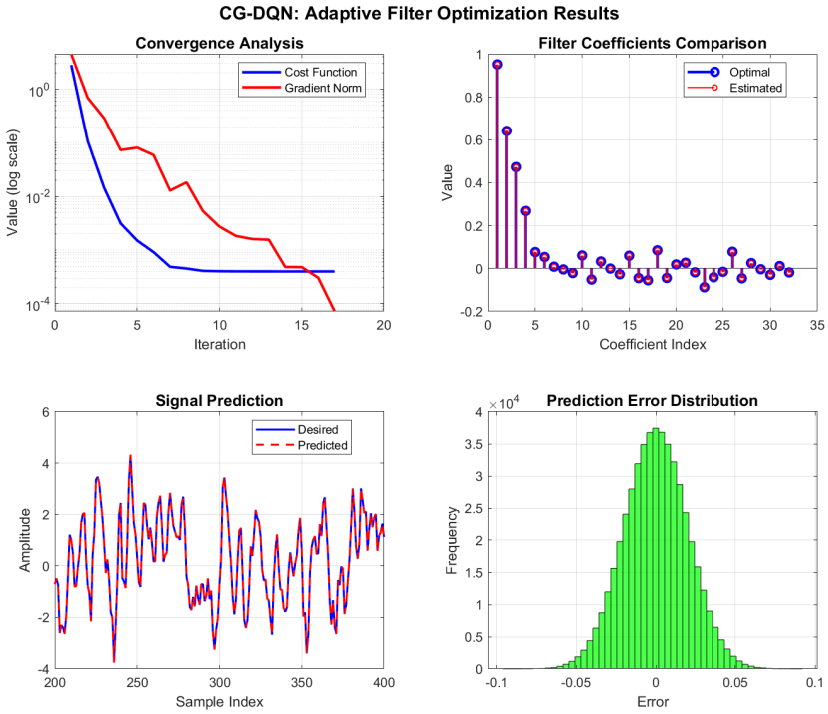


Figure 5: The results obtained from implementing the CG-DQN algorithm on the adaptive filter design problem.

$$\begin{aligned} \min \quad & \sum_{i=1}^{K} p_i \\ s.t. \quad & R_i(\mathbf{p}) \geq R_{\min}, \qquad i = 1, \ldots, K, \\ & p_i^{\min} \leq p_i \leq p_i^{\max}, \qquad i = 1, \ldots, K. \end{aligned} \tag{28}$$

By incorporating a penalty term for constraint violations, we reformulate the constrained problem into an unconstrained one.

$$\min_{\mathbf{p}} J(\mathbf{p}) = \sum_{i=1}^{K} p_i + \lambda \sum_{i=1}^{K} [\max(0, R_{\min} - R_i(\mathbf{p}))]^2. \tag{29}$$

For the bound constraints, we also employ a projection operator that maps the value of p_i to the closest point within the interval $[p_i^{\min}, p_i^{\max}]$.

This penalized problem was solved using the BFGS, CG-RMIL, and the proposed CG-DQN algorithms. The simulation parameters are summarized in Table 1.

Table 1: Simulation parameter settings

Parameter	**Value**
Number of users	$K = 5$
Minimum rate requirement	$R_{\min} = 1.0$ bps/Hz
Noise power	$\sigma^2 = 0.01$
Transmit power constraints	$0.001 \text{ W} \leq p_i \leq 2.0 \text{ W}$
Penalty parameter	$\lambda = 10$
Direct channel gains	$h_{ii} \sim \mathcal{U}(0.5, 1.0)$
Interference channel gains	$h_{ij} \sim \mathcal{U}(0.1, 0.3),\ i \neq j$

The results obtained from the BFGS, CG-RMIL, and CG-DQN algorithms are shown in Figure 6. The algorithms were programmed to stop running if $\|\nabla J(\mathbf{p})\| < 10^{-4}$. The approximate value of the ith component of the gradient is obtained from the following relation:

$$\nabla^i J(\mathbf{p}) = \frac{J(\mathbf{p} + 10^{-6} I_i) - J(\mathbf{p})}{10^{-6}},$$

where I_i is the ith column of the identity matrix.

In Figure 6, the first subplot illustrates the convergence behavior of the algorithms. As shown, the proposed CG-DQN method converges to a stationary point faster than CG-RMIL. The BFGS method, however, fails to converge. As one can see in this figure, the BFGS method showed good performance in the first 4 iterations and even achieved greater reduction compared to the other two methods. However, from the fifth iteration onward, the decreasing trend of the algorithm became very slow and almost stopped. The step length from the fourth iteration onward is very small, and the reason for this appears to be the approximate calculation of both the gradient and the Hessian. Although the direction calculated using the modified Cholesky decomposition is a descent direction, but from iteration 4 onward, the step length that can achieve sufficient reduction in the algorithm has become very small (this means that the approximations are not suitable), and practically no improvement is obtained in the algorithm.

Despite their different convergence speeds, CG-DQN and CG-RMIL both arrive at essentially the same optimal solution, which is verified in the remaining subplots.

The second subplot shows the optimized transmit power allocation. Both algorithms produce similar nonuniform distributions, reflecting the underlying channel and interference conditions: Users with weaker channels are assigned higher power to compensate, while those with stronger channels transmit at much lower levels.

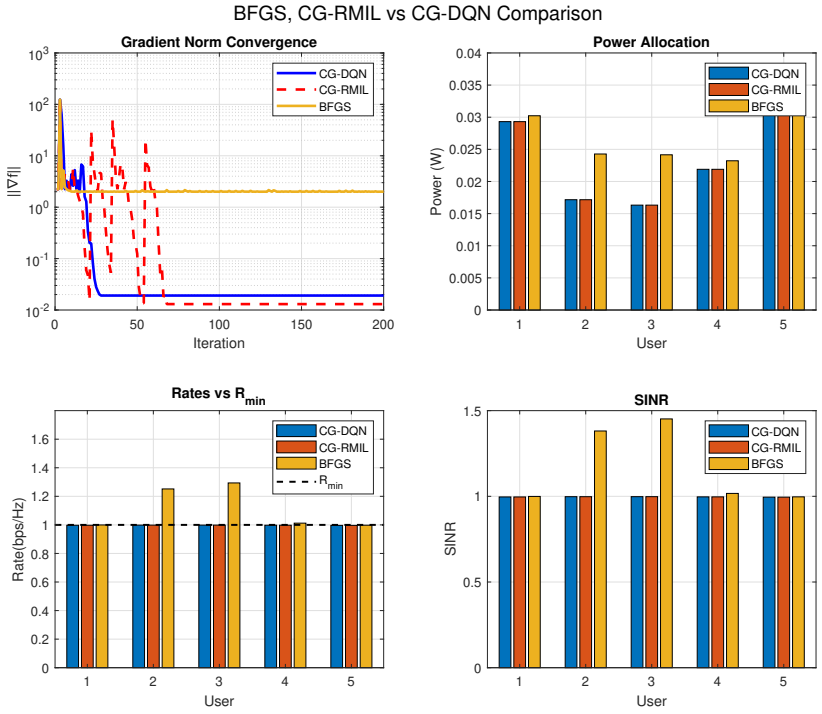


Figure 6: The results obtained from implementing the BFGS, CG-RMIL and CG-DQN algorithms on the power allocation problem.

The third subplot compares the achieved data rates with the minimum requirement. Both algorithms allocate just enough power to meet this threshold, confirming that the optimization avoids unnecessary power expenditure.

Finally, the fourth subplot shows the resulting signal-to-interference-plus-noise ratio (SINR) for each user. The SINR levels align with the allocated powers and achieved rates, showing that each method precisely adjusts the transmit powers to satisfy the performance requirement while minimizing total power consumption.

These results confirm that the proposed algorithm provides an efficient solution to the power allocation problem in wireless communication systems.

6 Conclusion

This paper presents a novel algorithmic framework for unconstrained optimization, introducing a hybrid method that innovatively combines the RMIL conjugate gradient approach with a diagonal quasi-Newton strategy. The algorithm is built on rigorous mathematical foundations, ensuring formally guaranteed global convergence properties. Extensive numerical experiments validate the computational efficiency of the proposed method. The evaluations, conducted on a diverse set of standard test problems, as well as two real-world engineering applications an adaptive filter design problem and a power allocation problem, demonstrate its practical superiority. The results indicate that the hybrid method not only outperforms the standard RMIL conjugate gradient method but also shows significant advantages over the classical BFGS algorithm in terms of computational time.

Funding

This research received no external funding.

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] Aini, N., Hajar, N., Mamat, M., Zull, N. and Rivaie, M., *Hybrid quasi-Newton and conjugate gradient method for solving unconstrained optimization problems*, J. Eng. Appl. Sci. 12(18) (2017), 4627–4631.

[2] Andrei, N., *An unconstrained optimization test functions collection*, Adv. Model. Optim. 10 (2008), 147–161.

[3] Andrei, N., *Accelerated conjugate gradient algorithm with finite difference Hessian/vector product approximation for unconstrained optimization*, J. Comput. Appl. Math. 230(2) (2009), 570–582.

[4] Broyden, C. G., *The convergence of a class of double-rank minimization algorithms. Part I. The new algorithm*, J. Inst. Math. Appl. 6 (1970), 76–90.

[5] Cao, H. and Li, W., *A diagonal quasi-Newton method based on automatic differentiation for nonlinear equations*, Calcolo 62(2) (2025), 1–17.

[6] Davidon, W. C., *Variable metric method for minimization*, Technical Report ANLC5990 (Revised), Argonne National Laboratory, Argonne, 1959.

[7] Dolan, E. D. and Moré, J. J., *Benchmarking optimization software with performance profiles*, Math. Program. 91 (2002), 201–213.

[8] Enshaei, S., Leong, W. H. and Farid, M., *Diagonal quasi-Newton method via variational principle under generalized Frobenius norm*, Optim. Methods Softw. 31(6) (2016), 1258–1271.

[9] Fletcher, R., *A new approach to variable metric algorithms*, Comput. J. 13(3) (1970), 317–322.

[10] Fletcher, R. and Powell, M. J. D., *A rapidly convergent descent method for minimization*, Comput. J. 6 (1963), 163–168.

[11] Goldfarb, A., *A family of variable metric methods derived by variational means*, Math. Comput. 24(109) (1970), 23–26.

[12] Haykin, S., *Adaptive filter theory*, Prentice-Hall, Inc., Upper Saddle River, NJ, 1996.

[13] Huynh, D. Q. and Hwang, F. N., *An accelerated structured quasi-Newton method with a diagonal second-order Hessian approximation for nonlinear least squares problems*, J. Comput. Appl. Math. 442 (2024), 115718.

[14] Liu, M., Zhang, J., Xiong, K., Zhang, M., Fan, P. and Letaief, K. B., *Effective user clustering and power control for multiantenna uplink NOMA transmission*, IEEE Trans. Wireless Commun. 21(11) (2022), 8995–9009.

[15] Monson, H., *Statistical digital signal processing and modeling*, John Wiley & Sons, Hoboken, NJ, 1996.

[16] Nocedal, J. and Wright, S. J., *Numerical optimization*, Springer Series in Operations Research and Financial Engineering, Springer, New York, 2006.

[17] Osinuga, I. and Stanimirovic, P., *Another hybrid BFGS-conjugate gradient algorithm for unconstrained optimization*, AIP Conf. Proc. 2852 (2023), 130006.

[18] Polyak, B. T., *The conjugate gradient method in extremal problems*, USSR Comput. Math. Math. Phys. 9(4) (1969), 94–112.

[19] Rivaie, M., Mamat, M., Leong, W. and Ismail, M., *A new conjugate gradient coefficient for large scale nonlinear unconstrained optimization*, Int. J. Math. Anal. 6(23) (2012), 1131–1146.

[20] Shanno, D.F.d, *Conditions of quasi-Newton methods for function minimization*, Math. Comput. 24(111) (1970), 647–650.

[21] Sim, H. S., Leong, W. J., Chen, C. Y. and Ibrahim, S. N. I., *Diagonal quasi-Newton updating formula using log-determinant norm*, in: *Advances in Industrial and Applied Mathematics: Proceedings of the 23rd Malaysian National Symposium of Mathematical Sciences (SKSM23)*, (2016), 1–7.

[22] Zhu, Z., Zhu, X. and Tan, Z., *An accelerated conjugate gradient method with adaptive two-parameter with applications in image restoration*, Comput. Appl. Math. 43 (2024), Article 116.

[23] Zola, E. and Martin-Escalona, I., *A robust user association, backhaul routing, and switching off model for a 5G network with variable traffic demands*, IEEE Access 8 (2020), 96714–96726.

[24] Zoutendijk, G., *Nonlinear programming, computational methods*, in: J. Abadie (ed.), *Integer and Nonlinear Programming*, North-Holland Publishing Company, Amsterdam and London, (1970), 37–86.

Iranian Journal of Numerical Analysis and Optimization
Vol. 17, No. 1, 2027, pp 380–415
https://doi.org/10.22067/ijnao.2026.100294.1934
https://ijnao.um.ac.ir/

CC BY NC

Research Article

Mathematical analysis of Mpox transmission using an SAIR-SI model with optimal control strategies

T. Janan, F. Fatmawati*, and A. Hasan

Abstract

This study presents an SAIR-SI (susceptible-asymptomatic-symptomatic-recovered-susceptible-infected) mathematical model of Mpox transmission using real data from the United States. First, the well-posedness, boundedness, and nonnegativity of the solutions are analyzed. Then, it is shown that the model admits three equilibria: Disease-free, human-endemic, and endemic. The local stability of the disease-free equilibrium is analyzed based on the Jacobian matrix, while the global stability of all equilibria is examined by defining distinct Lyapunov functions. Sensitivity analysis highlights transmission strength and progression rate as major determinants of epidemic behavior. Numerical simulations demonstrate that higher transmission

*Corresponding author

Received 6 August 2026; revised 9 September 2026; accepted 19 September 2026

Tuhfatul Janan
Department of Mathematics, Faculty of Science and Technology, Universitas Airlangga, Surabaya, Indonesia. e-mail: tuhfatuljanan1@gmail.com. Department of Mathematics Education, Faculty of Education and Teacher Training, Institut Ahmad Dahlan Probolinggo, Probolinggo, Indonesia

Fatmawati Fatmawati
Department of Mathematics, Faculty of Science and Technology, Universitas Airlangga, Surabaya, Indonesia. e-mail: fatmawati@fst.unair.ac.id

Agus Hasan
Department of ICT and Natural Sciences, Norwegian University of Science and Technology, Alesund, Norway. e-mail: agus.hasan@ntnu.no

How to cite this article
Janan, T., Fatmawati, F. and Hasan, A., Mathematical analysis of Mpox transmission using an SAIR-SI model with optimal control strategies. *Iran. J. Numer. Anal. Optim.*, 2027; 17(1): 380-415. https://doi.org/10.22067/ijnao.2026.100294.1934

accelerates spread, while slower progression from asymptomatic to symptomatic stages prolongs infectiousness and increases disease burden. Furthermore, two control measures, namely vaccination and rodenticide usage, are considered, and the optimal control problem is formulated based on Pontryagin's principle. Numerical simulations incorporating control measures demonstrate that the strategy combining both controls is the most effective approach for reducing the asymptomatic human, symptomatic human, and infected rodent subpopulations. The cost-effectiveness analysis also confirms that the combined implementation of vaccination and rodenticide usage is the most efficient strategy.

AMS subject classifications (2020): 37N25, 49K40, 92-10, 92B05, 93-10.

Keywords: Mpox disease: Asymptomatic humans; SAIR-SI model; Stability analysis; Optimal control.

1 Introduction

Monkeypox (Mpox) is a zoonotic disease caused by the Mpox virus, a member of the Orthopoxvirus genus, which is phylogenetically related to the variola virus [10]. The virus is classified into two major clades, namely Clade I (subclades Ia and Ib) and Clade II (subclades IIa and IIb). The first case was identified in monkeys in Denmark in 1958, while human infection was first reported in the Democratic Republic of Congo in 1970 [35]. Although Mpox remains endemic in several African regions, since May 2022 its transmission has expanded to nonendemic countries, primarily linked to international travel to Europe and North America [34]. With the rapid global increase in cases, the World Health Organization (WHO) declared Mpox a Public Health Emergency of International Concern on August 14, 2024, after more than 100,000 cases and 220 deaths had been reported across 120 countries [37].

Transmission occurs through direct contact with infected individuals, animal reservoirs, or contaminated objects [12]. Routes of infection include respiratory droplets, skin lesions, body fluids [28], and possible vertical transmission from mother to fetus or newborn [38]. In zoonotic contexts, transmission may occur through bites or contact with fluids of infected animals, such as during hunting or consumption of bushmeat [35, 12, 2]. Fomite transmission via contaminated materials such as clothing and bedding has also been documented [12]. The incubation period of Mpox is generally from 6 to 13 days but can range from 5 to 21 days [30, 31]. During this period, infected individuals remain asymptomatic and clinically well. Typical clinical manifestations include rash, fever, sore throat, headache, myalgia, back pain, fatigue, and lymphadenopathy. Most symptoms will last for approximately 2–4 weeks. With good supportive care, the vast

majority of cases will improve. Severe outcomes and mortality are most likely in newborns, children, pregnant women, and immunocompromised patients, including patients in the advanced stages of HIV infection. According to the WHO, the case fatality ratio ranges from 0.1% to 10% [36].

With several hundred cases developing daily, the United States (U.S.) is currently one of the countries most affected by the recent Mpox outbreaks, which occurred between July and September 2022 [13]. The Department of Health and Human Services distributed over one million vials of the JYNNEOS vaccine nationwide from August 2022 to August 2024, in an effort to reduce transmission and the severity of this disease [11]. Governmental efforts have also aimed at promoting research into Mpox, concentrating on viral mechanisms for transmission, host immune responses, and optimization of vaccine dosing regimens to increase supply and facilitate new vaccine development. In addition to vaccination strategies, therapeutic interventions involving the use of antivirals and monoclonal antibodies, along with laboratory diagnostic system strengthening, remain crucial for patient management and epidemiological surveillance [32].

Mathematical modeling has been consistently used as a tool for understanding infectious disease dynamics and formulating control strategies effectively [26, 27]. The classical compartmental model, which classifies the population into susceptible, infected, and recovered subpopulations, is the bedrock of other more developed epidemiological models [15, 18, 40]. In recent times, there has been the inclusion of extra compartments such as vaccinated [3, 33] and asymptomatic [1, 29] status to give these models real-world representation in public health contexts. This extension has greatly improved the ability of the models to assess the effectiveness of intervention measures. For example, Al-Shomrani, Musa, and Yusuf [7] illustrated that the inclusion of extra subpopulations leads to more accurate estimation of policy impacts.

In the case of Mpox, the transmission models generally consider interactions between the human and rodent populations. Thus, Alshehri and Ullah [6] presented a framework that divided the human subpopulations into susceptible, infected, quarantined, and recovered groups while dividing the rodent population into two classes of susceptible and infected populations. Based on this, the extension introduced by Allehiany et al. [4] further extended the classes by incorporating exposed compartments for humans and rodents. Further refinements were made by Wireko et al. [39], who integrated an isolation compartment into the human model creating six categories: Susceptible, exposed, infected, isolated, quarantined, and recovered. In the refinement to the rodents, Yapışkan et al. [41] incorporated a recovered subpopulation into the standard susceptible-infected structure, while Al-Shomrani, Musa, and Yusuf [7] further extended this by incorporating an exposed class.

The WHO also emphasized that asymptomatic individuals with Mpox virus infection can still transmit the disease in the absence of symptoms and clinical signs [35]. Li et al. [24] have ex-

plicitly modeled asymptomatic individuals to realistically capture the dynamics of transmission. According to their analysis, the asymptomatic group is crucial in sustaining chains of infection. This calls for prioritizing asymptomatic carriers with intervention measures such as vaccination and personal protection. Likewise, Fernandes and Maldonado [17] demonstrated that asymptomatic carriers can be capable of spread the virus via sexual contact, hence their relevance in determining the probabilities of transmission. These findings are of great importance to public health authorities in developing prevention policies, particularly those dealing with sexually transmitted infections with latent or asymptomatic routes of transmission. Additional studies related to Mpox modeling can be found in [8, 20, 22, 25, 42].

It is against this backdrop that this study presents an SAIR-SI mathematical model for Mpox transmission in the United States, which explicitly includes asymptomatic humans. In developing the model, recent evidence has been used to highlight the critical role played by asymptomatic individuals in disease transmission, which previous models have often ignored. The proposed model is subsequently extended by incorporating an optimal control framework to further investigate effective control strategies for suppressing the spread of Mpox. In this extended model, two control measures are considered, namely vaccination and rodenticide usage, which are evaluated using Pontryagin's principle. Each control strategy is systematically illustrated through comparison with the uncontrolled scenario in order to assess its effectiveness in reducing the asymptomatic human, symptomatic human, and infected rodent subpopulations. Overall, compared with previous Mpox transmission models that incorporate asymptomatic individuals, this study extends the existing framework by using data from the United States in 2022, considering vaccination and rodenticide usage as control measures, and evaluating their cost-effectiveness using the ICER method, as summarized in Table 1.

Table 1: Comparison of recent Mpox transmission models

Reference	Data	Control measures	Cost-effectiveness
[7]	Nigeria, 2022	None	None
[17]	Multiple countries	None	None
[24]	Nigeria, 2022-2023	Personal protection, vaccination, and reduced animal contact	None
This study	United States, 2022	Vaccination and rodenticide usage	ICER method

2 Mathematical model and basic properties

2.1 Mathematical model

The SAIR-SI mathematical model explicitly captures the transmission dynamics of Mpox, incorporating interactions between human and rodent populations. The human host population is stratified into four epidemiological compartments: susceptible humans (S_h), asymptomatic humans (A_h), symptomatic infected humans (I_h), and recovered humans (R_h). In contrast, the rodent population is structured into two classes: Susceptible rodents (S_r) and infected rodents (I_r). The total human population at any time is expressed as $N_h = S_h + A_h + I_h + R_h$, while the total rodent population is represented as $N_r = S_r + I_r$. The transmission diagram of Mpox dynamics is illustrated in Figure 1.

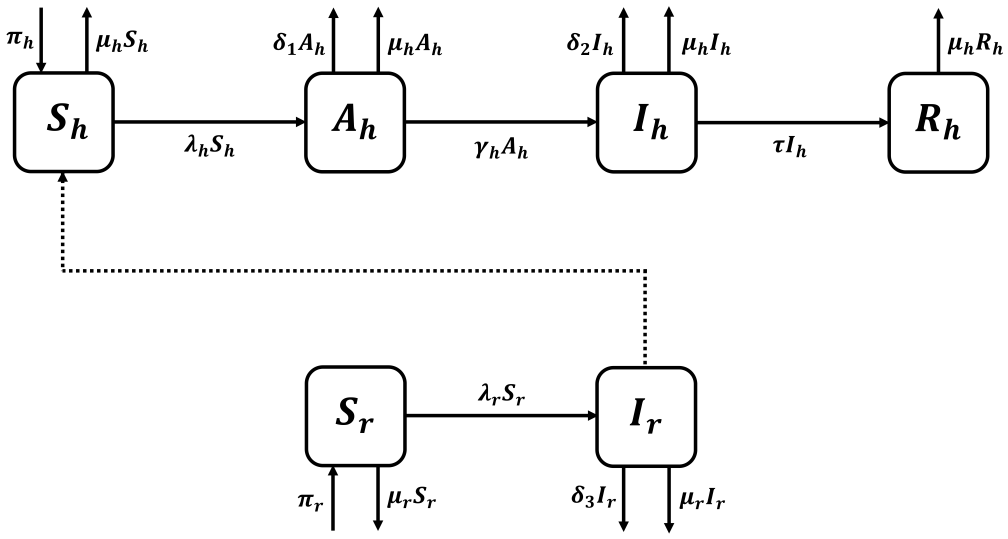


Figure 1: Mpox transmission diagram

Within the human host population, the pool of susceptible humans grows through recruitment at a constant rate (π_h). These individuals are subject to infection either through effective contact with asymptomatic humans, symptomatic infected humans, or infectious rodents, and may also be lost due to natural mortality at rate μ_h. The risk of infection in humans is characterized by the force of infection, $\lambda_h = \frac{\beta_1 A_h + \beta_2 I_h + \beta_3 I_r}{N_h}$, where β_1, β_2, and β_3 represent the respective effective contact rates with asymptomatic humans, symptomatic humans, and infected rodents.

Asymptomatic humans, despite not showing clinical manifestations, remain infectious and thus contribute to onward transmission. They may transition to the symptomatic class at rate γ_h or be removed from the population due to Mpox-induced mortality (δ_1) or natural causes (μ_h). Symptomatic individuals can either recover at rate τ or experience disease-induced death (δ_2) and natural death (μ_h). In the rodent reservoir population, susceptible rodents enter the system at a recruitment rate (π_r) and are exposed to infection through effective contact with infectious rodents, governed by the force of infection $\lambda_r = \frac{\beta_4 I_r}{N_r}$, with β_4 denoting the effective contact rate. Both susceptible and infected rodents are subject to natural mortality (μ_r), whereas infected rodents additionally face Mpox-induced mortality at rate δ_3. A detailed description of each parameter is provided in Table 2.

Table 2: Description of parameters

Parameter	Description
π_h	Rate of entry of susceptible humans
π_r	Rate of entry of susceptible rodents
β_1	Transmission rate via contact with asymptomatic humans
β_2	Transmission rate via contact with symptomatic humans
β_3	Transmission rate from infected rodents to humans
β_4	Transmission rate among rodents
γ_h	Rate of progression from asymptomatic to symptomatic infection in humans
τ	Recovery rate of symptomatic humans
δ_1	Mpox-induced death rate in asymptomatic humans
δ_2	Mpox-induced death rate in symptomatic humans
δ_3	Mpox-induced death rate in rodents
μ_h	Natural death rate of humans
μ_r	Natural death rate of rodents

The dynamics of Mpox transmission are mathematically described using the following system of differential equations:

$$\begin{aligned}
\frac{dS_h}{dt} &= \pi_h - (\lambda_h + \mu_h)S_h, \\
\frac{dA_h}{dt} &= \lambda_h S_h - (\gamma_h + \delta_1 + \mu_h)A_h, \\
\frac{dI_h}{dt} &= \gamma_h A_h - (\tau + \delta_2 + \mu_h)I_h, \\
\frac{dR_h}{dt} &= \tau I_h - \mu_h R_h, \\
\frac{dS_r}{dt} &= \pi_r - (\lambda_r + \mu_r)S_r, \\
\frac{dI_r}{dt} &= \lambda_r S_r - (\delta_3 + \mu_r)I_r,
\end{aligned} \tag{1}$$

with the initial conditions $S_h(0) > 0$, $A_h(0) \geq 0$, $I_h(0) \geq 0$, $R_h(0) \geq 0$, $S_r(0) > 0$, and $I_r(0) \geq 0$.

2.2 Well-posedness analysis

We begin by establishing the well-posedness of model (1) through the application of the contraction mapping theorem, aligning with the methodological framework presented by Elsonbaty et al. [16]. A model is said to be well-posed if its solution satisfies three key criteria: Existence, uniqueness, and continuous dependence on initial conditions (stability). We consider the first differential equation of model (1), expressed as

$$\frac{dS_h}{dt} = \pi_h - (\lambda_h + \mu_h)S_h.$$

Integrating both sides yields the corresponding integral representation:

$$S_h(t) = S_h(0) + \int_0^t (\pi_h - (\lambda_h + \mu_h)S_h)\, ds.$$

Define the operator T acting on S_h by

$$T(S_h(t)) = S_h(0) + \int_0^t (\pi_h - (\lambda_h + \mu_h)S_h)\, ds.$$

Let Y denote a complete metric space defined as

$$Y = \{S_h : [0, a] \to \mathbb{R} \mid S_h \text{ is continuous}\},$$

with the supremum norm,

$$d(S_{h1}, S_{h2}) = \|S_{h1} - S_{h2}\|_\infty = \sup_{t\in[0,a]} |S_{h1}(t) - S_{h2}(t)|.$$

For any $S_{h_1}, S_{h_2} \in Y$, the following inequality holds:

$$\begin{aligned} \left|T(S_{h_1}(t)) - T(S_{h_2}(t))\right| &= \left|\int_0^t -(\lambda_h + \mu_h)\,(S_{h_1} - S_{h_2})\, ds\right| \\ &\leq Ct\,\|S_{h_1} - S_{h_2}\|_\infty, \end{aligned}$$

where $C = \sup_{s\in[0,a]} (\lambda_h + \mu_h)$.
Thus,

$$\|T(S_{h1}) - T(S_{h2})\|_\infty \leq Ca\|S_{h1} - S_{h2}\|_\infty. \tag{2}$$

Therefore, the operator T is a contraction provided that $Ca < 1$, or equivalently $a < \frac{1}{C}$. According to the contraction mapping principle, under this condition there exists a unique fixed point $S_h(t) \in Y$ such that $S_h(t) = T(S_h(t))$.

To investigate the stability of the solution, consider two solutions $S_{h1}(t)$ and $S_{h2}(t)$ with distinct initial values. The corresponding integral forms are

$$S_{h1}(t) = S_{h1}(0) + \int_0^t (\pi_h - (\lambda_h + \mu_h)S_{h1}(s))\, ds,$$
$$S_{h2}(t) = S_{h2}(0) + \int_0^t (\pi_h - (\lambda_h + \mu_h)S_{h2}(s))\, ds.$$

Subtracting and estimating the difference yields

$$\|S_{h_1}(t) - S_{h_2}(t)\|_\infty \leq \|S_{h_1}(0) - S_{h_2}(0)\|_\infty + Ca\, \|S_{h_1} - S_{h_2}\|_\infty,$$

or equivalently,

$$(1 - Ca)\, \|S_{h_1}(t) - S_{h_2}(t)\|_\infty \leq \|S_{h_1}(0) - S_{h_2}(0)\|_\infty.$$

Assuming $0 < \|S_{h1}(0) - S_{h2}(0)\|_\infty \leq \theta$, it follows that

$$0 < \|S_{h1}(t) - S_{h2}(t)\|_\infty \leq \frac{\theta}{1 - Ca} = \varepsilon, \tag{3}$$

where $Ca < 1$ and $C = \sup_{s\in[0,a]}(\lambda_h + \mu_h)$.
This confirms the stability of the solution $S_h(t) \in Y$.

Following similar reasoning, one can show that all other state variables, namely $A_h(t)$, $I_h(t)$, $R_h(t)$, $S_r(t)$, and $I_r(t)$, also satisfy the conditions for existence, uniqueness, and stability. Hence, we conclude that model (1) is well-posed.

2.3 Boundedness and nonnegativity

We now investigate the boundedness of the model's solutions. For the human population, we have

$$\frac{dN_h}{dt} = \pi_h - \mu_h N_h - \delta_1 A_h - \delta_2 I_h \leq \pi_h - \mu_h N_h. \tag{4}$$

Solving this differential inequality leads to

$$N_h(t) \leq \frac{\pi_h}{\mu_h} + \left(N_h(0) - \frac{\pi_h}{\mu_h}\right) e^{-\mu_h t}.$$

Thus,

$$\lim_{t\to\infty} N_h(t) \leq \frac{\pi_h}{\mu_h}. \tag{5}$$

Similarly, for the rodent population, we get

$$\frac{dN_r}{dt} = \pi_r - \mu_r N_r - \delta_3 I_r \leq \pi_r - \mu_r N_r. \tag{6}$$

Solving this differential inequality yields

$$N_r(t) \leq \frac{\pi_r}{\mu_r} + \left(N_r(0) - \frac{\pi_r}{\mu_r}\right) e^{-\mu_r t},$$

implying

$$\lim_{t\to\infty} N_r(t) \leq \frac{\pi_r}{\mu_r}. \tag{7}$$

Hence, all solutions remain bounded within the biologically feasible region:

$$\Omega = \left\{(S_h, A_h, I_h, R_h, S_r, I_r) \in \mathbb{R}^6_+ \mid N_h \leq \frac{\pi_h}{\mu_h},\ N_r \leq \frac{\pi_r}{\mu_r}\right\}. \tag{8}$$

Next, we proceed to verify the nonnegativity of the solutions of model (1), as formally stated in Theorem 1.

Theorem 1. Let $S_h(t), A_h(t), I_h(t), R_h(t), S_r(t)$, and $I_r(t)$ denote the solutions of model (1). If the initial values are nonnegative, then each solution remains nonnegative for all $t \geq 0$.

Proof. Consider the first differential equation of model (1):

$$\frac{dS_h}{dt} = \pi_h - (\lambda_h + \mu_h) S_h.$$

Multiplying both sides by the integrating factor $e^{\mu_h t + \int_0^t \lambda_h\, ds}$ yields

$$\frac{d}{dt}\left(e^{\mu_h t + \int_0^t \lambda_h\, ds} S_h\right) = \pi_h e^{\mu_h t + \int_0^t \lambda_h\, ds}.$$

Solving this equation, we obtain

$$S_h(t) = \left(\int_0^t \pi_h e^{\mu_h y + \int_0^y \lambda_h \, ds} dy + S_h(0) \right) e^{-\mu_h t - \int_0^t \lambda_h \, ds} \geq 0. \tag{9}$$

Since both the exponential term and the integrand are nonnegative, and $S_h(0) \geq 0$, it follows that $S_h(t) \geq 0$ for all $t \geq 0$. Applying analogous steps to the remaining differential equations confirms that the other state variables $A_h(t), I_h(t), R_h(t), S_r(t), I_r(t)$ also remain nonnegative for all $t \geq 0$, provided their respective initial conditions are nonnegative. □

3 Equilibrium and stability analysis

3.1 Equilibrium and basic reproduction number

In this section, we investigate the equilibrium states of model (1) by setting the time derivatives of all state variables to zero, as outlined in [9]. This yields the steady-state conditions

$$\frac{dS_h}{dt} = \frac{dA_h}{dt} = \frac{dI_h}{dt} = \frac{dR_h}{dt} = \frac{dS_r}{dt} = \frac{dI_r}{dt} = 0.$$

Under these conditions, model (1) reduces to the following system of algebraic equations:

$$\begin{aligned} \pi_h - (\lambda_h + \mu_h) S_h &= 0, \\ \lambda_h S_h - m_1 A_h &= 0, \\ \gamma_h A_h - m_2 I_h &= 0, \\ \tau I_h - \mu_h R_h &= 0, \\ \pi_r - (\lambda_r + \mu_r) S_r &= 0, \\ \lambda_r S_r - m_3 I_r &= 0, \end{aligned} \tag{10}$$

where m_1, m_2, m_3 are defined as $m_1 = \gamma_h + \delta_1 + \mu_h, m_2 = \tau + \delta_2 + \mu_h$, and $m_3 = \delta_3 + \mu_r$. The disease-free equilibrium (DFE) corresponds to a scenario in which there is no active Mpox infection in either the human or rodent populations, that is, $A_h = I_h = I_r = 0$. Thus, the DFE of model (1) is given by

$$X^0 = (S_h^0, A_h^0, I_h^0, R_h^0, S_r^0, I_r^0) = \left(\frac{\pi_h}{\mu_h}, 0, 0, 0, \frac{\pi_r}{\mu_r}, 0 \right). \tag{11}$$

We now proceed to derive the basic reproduction number, denoted by $\mathcal{R}_0$, which quantifies the average number of secondary infections generated by a single infectious individual in a fully

susceptible population. In this study, $\mathcal{R}_0$ is computed using the next-generation matrix (NGM) approach. The infected compartments in model (1) are A_h, I_h, and I_r. Focusing on their dynamics, the infected subsystem can be written as

$$\begin{pmatrix} \frac{dA_h}{dt} \\ \frac{dI_h}{dt} \\ \frac{dI_r}{dt} \end{pmatrix} = \begin{pmatrix} \lambda_h S_h \\ 0 \\ \lambda_r S_r \end{pmatrix} - \begin{pmatrix} m_1 A_h \\ -\gamma_h A_h + m_2 I_h \\ m_3 I_r \end{pmatrix}. \tag{12}$$

Linearizing system (12) around the DFE leads to the construction of the Jacobian matrices F and V, representing new infection terms and transition dynamics, respectively:

$$F = \begin{pmatrix} \frac{(\beta_1-\lambda_h)S_h}{N_h} & \frac{(\beta_2-\lambda_h)S_h}{N_h} & \frac{\beta_3 S_h}{N_h} \\ 0 & 0 & 0 \\ 0 & 0 & \frac{(\beta_4-\lambda_r)S_r}{N_r} \end{pmatrix}, \qquad V = \begin{pmatrix} m_1 & 0 & 0 \\ -\gamma_h & m_2 & 0 \\ 0 & 0 & m_3 \end{pmatrix}.$$

Evaluating matrices F and V at the DFE (11), we get

$$\mathbb{F} = \begin{pmatrix} \beta_1 & \beta_2 & \beta_3 \\ 0 & 0 & 0 \\ 0 & 0 & \beta_4 \end{pmatrix}, \qquad \mathbb{V} = \begin{pmatrix} m_1 & 0 & 0 \\ -\gamma_h & m_2 & 0 \\ 0 & 0 & m_3 \end{pmatrix}.$$

The NGM is then computed as

$$K = \mathbb{F}\mathbb{V}^{-1} = \begin{pmatrix} \frac{\beta_1 m_2 + \beta_2 \gamma_h}{m_1 m_2} & \frac{\beta_2}{m_2} & \frac{\beta_3}{m_3} \\ 0 & 0 & 0 \\ 0 & 0 & \frac{\beta_4}{m_3} \end{pmatrix}.$$

The eigenvalues of matrix K are

$$\zeta_1 = 0, \quad \zeta_2 = \frac{\beta_1 m_2 + \beta_2 \gamma_h}{m_1 m_2}, \quad \text{and } \zeta_3 = \frac{\beta_4}{m_3}.$$

Accordingly, the basic reproduction number is given by

$$\mathcal{R}_0 = \max\left\{\mathcal{R}_0^h, \mathcal{R}_0^r\right\} = \max\left\{\frac{\beta_1 m_2 + \beta_2 \gamma_h}{m_1 m_2}, \frac{\beta_4}{m_3}\right\}. \tag{13}$$

Next, we derive the endemic equilibria of model (1), which represent the long-term persistence of Mpox within the population. Model (1) admits two distinct endemic equilibrium states: The

human-endemic equilibrium (HEE) and the endemic equilibrium (EE). The HEE corresponds to a scenario in which Mpox persists solely in the human population while the rodent population remains uninfected, that is, $A_h \neq 0$, $I_h \neq 0$, but $I_r = 0$. Under these conditions, the HEE is denoted by

$$X^+ = (S_h^+, A_h^+, I_h^+, R_h^+, S_r^+, I_r^+), \tag{14}$$

with component expressions

$$S_h^+ = \frac{\pi_h}{\lambda_h^+ + \mu_h}, \quad A_h^+ = \frac{\pi_h \lambda_h^+}{(\lambda_h^+ + \mu_h)m_1}, \quad I_h^+ = \frac{\pi_h \gamma_h \lambda_h^+}{(\lambda_h^+ + \mu_h)m_1 m_2},$$
$$R_h^+ = \frac{\pi_h \gamma_h \lambda_h^+ \tau}{(\lambda_h^+ + \mu_h)m_1 m_2 \mu_h}, \quad S_r^+ = \frac{\pi_r}{\mu_r}, \quad I_r^+ = 0,$$

where the infection rate λ_h^+ is given by

$$\lambda_h^+ = \frac{m_1 m_2 \mu_h (\mathcal{R}_0^h - 1)}{m_2 \mu_h + \gamma_h \mu_h + \gamma_h \tau}.$$

It follows that the HEE exists if and only if $\mathcal{R}_0^h > 1$, indicating that the basic reproduction number for humans must exceed unity for sustained transmission in the human population.

In contrast, the EE describes a state in which both humans and rodents are endemically infected, that is, $A_h \neq 0$, $I_h \neq 0$, and $I_r \neq 0$. The EE is denoted by

$$X^* = (S_h^*, A_h^*, I_h^*, R_h^*, S_r^*, I_r^*), \tag{15}$$

with component expressions

$$S_h^* = \frac{\pi_h}{\lambda_h^* + \mu_h}, \quad A_h^* = \frac{\pi_h \lambda_h^*}{(\lambda_h^* + \mu_h)m_1}, \quad I_h^* = \frac{\pi_h \gamma_h \lambda_h^*}{(\lambda_h^* + \mu_h)m_1 m_2},$$
$$R_h^* = \frac{\pi_h \gamma_h \lambda_h^* \tau}{(\lambda_h^* + \mu_h)m_1 m_2 \mu_h}, \quad S_r^* = \frac{\pi_r}{\lambda_r^* + \mu_r}, \quad I_r^* = \frac{\pi_r \lambda_r^*}{(\lambda_r^* + \mu_r)m_3},$$

where the infection rate for rodents is given by

$$\lambda_r^* = m_3(\mathcal{R}_0^r - 1),$$

and the infection rate λ_h^* for humans is determined as the positive root of the quadratic equation

$$a{\lambda_h^*}^2 + b\lambda_h^* + c = 0,$$

where the coefficients a, b, and c are constants with respect to λ_h^* and are given by

$$a = \pi_h m_3(\lambda_r^* + \mu_r)(m_2\mu_h + \gamma_h\mu_h + \gamma_h\tau),$$
$$b = \pi_h m_1 m_2 m_3 \mu_h(\lambda_r^* + \mu_r)(1 - \mathcal{R}_0^h) - \beta_3\pi_r\lambda_r^*\mu_h m_1 m_2,$$
$$c = -\beta_3\pi_r\lambda_r^*\mu_h^2 m_1 m_2.$$

When $\mathcal{R}_0^r = 1$, the equilibrium reduces to $X^* = X^+$. Therefore, the condition $\mathcal{R}_0^r > 1$ is required to ensure that $\lambda_r^* > 0$. Given that all model parameters are strictly positive, this directly implies that $a > 0$ and $c < 0$. This guarantees the existence of a unique positive solution for λ_h^*. Therefore, the existence and uniqueness of the EE are ensured if and only if $\mathcal{R}_0^r > 1$.

3.2 Stability of disease-free equilibrium

In this section, we investigate both the local and global asymptotic stability of the DFE of model (1). The local stability analysis is carried out via linearization of the system around the DFE, using the Jacobian matrix, as stated in Theorem 2.

Theorem 2. The DFE of model (1) is locally asymptotically stable if and only if $\mathcal{R}_0 < 1$.

Proof. The Jacobian matrix associated with model (1) is expressed as

$$J = \begin{pmatrix} A - \lambda_h - \mu_h & B_1 & B_2 & A & A & C \\ -A + \lambda_h & -B_1 - m_1 & -B_2 & -A & -A & -C \\ 0 & \gamma_h & -m_2 & 0 & 0 & 0 \\ 0 & 0 & \tau & -\mu_h & 0 & 0 \\ 0 & 0 & 0 & 0 & D - \lambda_r - \mu_r & E \\ 0 & 0 & 0 & 0 & -D + \lambda_r & -E - m_3 \end{pmatrix}, \tag{16}$$

with $A = \frac{\lambda_h S_h}{N_h}, B_i = \frac{(\lambda_h - \beta_i)S_h}{N_h}, C = \frac{-\beta_3 S_h}{N_h}, D = \frac{\lambda_r S_r}{N_r}, E = \frac{(\lambda_r - \beta_4)S_r}{N_r}$, and m_1, m_2, m_3 as defined in equation (10).
Evaluating the Jacobian matrix at the DFE (11), we obtain

$$J(X^0) = \begin{pmatrix} -\mu_h & -\beta_1 & -\beta_2 & 0 & 0 & -\beta_3 \\ 0 & \beta_1 - m_1 & \beta_2 & 0 & 0 & \beta_3 \\ 0 & \gamma_h & -m_2 & 0 & 0 & 0 \\ 0 & 0 & \tau & -\mu_h & 0 & 0 \\ 0 & 0 & 0 & 0 & -\mu_r & -\beta_4 \\ 0 & 0 & 0 & 0 & 0 & \beta_4 - m_3 \end{pmatrix}.$$

The characteristic equation corresponding to matrix $J(X^0)$ is given by

$$(\mu_h + \zeta)^2(\mu_r + \zeta)(-\beta_4 + m_3 + \zeta)(\zeta^2 + k_1\zeta + k_2) = 0, \tag{17}$$

where

$$k_1 = \frac{m_2^2 + \beta_2\gamma_h + m_1m_2(1 - \mathcal{R}_0^h)}{m_2} \text{ and } k_2 = m_1m_2(1 - \mathcal{R}_0^h).$$

This yields the eigenvalues $\zeta_1 = \zeta_2 = -\mu_h, \zeta_3 = -\mu_r, \zeta_4 = m_3(\mathcal{R}_0^r - 1)$, along with the roots of the quadratic equation

$$\zeta^2 + k_1\zeta + k_2 = 0. \tag{18}$$

It is evident that $\zeta_1, \zeta_2, \zeta_3 < 0$, and that $\zeta_4 < 0$ if and only if $\mathcal{R}_0^r < 1$. Furthermore, by applying the Routh–Hurwitz criterion, the roots of (18) possess negative real parts if and only if $k_1 > 0$ and $k_2 > 0$, which holds when $\mathcal{R}_0^h < 1$. Thus, the DFE is locally asymptotically stable if and only if $\mathcal{R}_0 < 1$. □

To assess the global asymptotic stability of the DFE, we construct a Lyapunov function, as formulated in Theorem 3.

Theorem 3. The DFE of model (1) is globally asymptotically stable if and only if $\mathcal{R}_0 < 1$.

Proof. Consider the Lyapunov function $L : D \to \mathbb{R}$ defined by

$$L = c_1A_h + c_2I_h + c_3I_r, \tag{19}$$

where $c_1, c_2, c_3 > 0$.
At the DFE, it holds that $L(X^0) = 0$, while $L > 0$ for all other states in the domain. Differentiating L with respect to time yields

$$\frac{\partial L}{\partial t} \leq (c_1(\beta_1 - m_1) + c_2\gamma_h)\, A_h + (c_1\beta_2 - c_2m_2)\, I_h + (c_1\beta_3 + c_3(\beta_4 - m_3))\, I_r. \tag{20}$$

To remove the coefficient of A_h, we set

$$c_1 = \frac{m_2\gamma_h}{\beta_2\gamma_h + m_1m_2(1-\mathcal{R}_0^h)}, \quad c_2 = 1,$$

under the condition $\mathcal{R}_0^h < 1$.
Next, we choose

$$c_3 = \frac{\beta_3 m_2}{\beta_2 m_3(1-\mathcal{R}_0^r)},$$

under the condition $\mathcal{R}_0^r < 1$.
Substituting c_1, c_2, and c_3 into inequality (20) gives

$$\frac{\partial L}{\partial t} \leq -\frac{m_1m_2^2(1-\mathcal{R}_0^h)}{\beta_2\gamma_h + m_1m_2(1-\mathcal{R}_0^h)} I_h - \frac{\beta_3 m_1 m_2^2(1-\mathcal{R}_0^h)}{\beta_2^2\gamma_h + \beta_2 m_1 m_2(1-\mathcal{R}_0^h)} I_r. \tag{21}$$

Since $\mathcal{R}_0^h < 1$, it can be concluded that $\frac{\partial L}{\partial t} < 0$ for all $X \neq X^0$, and $\frac{\partial L}{\partial t} = 0$ only when the state lies at the DFE. Therefore, by LaSalle's invariance principle [21], the DFE is globally asymptotically stable if and only if $\mathcal{R}_0 < 1$. □

3.3 Global stability of endemic equilibrium

We now proceed to analyze the global asymptotic stability of the HEE of model (1), employing a Lyapunov function approach, as established in Theorem 4.

Theorem 4. Let the HEE exist, which occurs when $\mathcal{R}_0^h > 1$. The HEE of model (1) is globally asymptotically stable if and only if $\mathcal{R}_0^r < 1$.

Proof. Consider the Lyapunov function $L : D \to \mathbb{R}$ defined by

$$\begin{aligned} L = &\left(S_h - S_h^+ - S_h^+ \ln \frac{S_h}{S_h^+}\right) + \left(A_h - A_h^+ - A_h^+ \ln \frac{A_h}{A_h^+}\right) \\ &+ c_1\left(I_h - I_h^+ - I_h^+ \ln \frac{I_h}{I_h^+}\right) + c_2\left(S_r - S_r^+ - S_r^+ \ln \frac{S_r}{S_r^+}\right) + c_3 I_r, \end{aligned} \tag{22}$$

where $c_1 = \frac{\beta_2 I_h^+ S_h^+}{\gamma_h A_h^+ N_h^+}$, $c_2 = \frac{S_r^+}{\beta_4}$, and $c_3 = \frac{S_r^+}{m_3(1-\mathcal{R}_0^r)}$ under the condition $\mathcal{R}_0^r < 1$.
Denoting logarithmic expressions by L_1 through L_4, it is evident that $L(X^+) = L_1^+ + L_2^+ + c_1 L_3^+ + c_2 L_4^+ + c_3 I_r^+ = 0$, while $L > 0$ for all $X \neq X^+$, due to the convex nature of each component. This

ensures that the Lyapunov function is positive definite with respect to the HEE. Differentiating L with respect to time yields

$$\begin{aligned}\frac{\partial L}{\partial t} =& \left(1-\frac{S_h^+}{S_h}\right)\frac{dS_h}{dt} + \left(1-\frac{A_h^+}{A_h}\right)\frac{dA_h}{dt} + c_1\left(1-\frac{I_h^+}{I_h}\right)\frac{dI_h}{dt} \\ &+ c_2\left(1-\frac{S_r^+}{S_r}\right)\frac{dS_r}{dt} + c_3\frac{dI_r}{dt}. \end{aligned} \tag{23}$$

Substituting model (1) into equation (23), we get

$$\begin{aligned}\frac{\partial L}{\partial t} =& \left(1-\frac{S_h^+}{S_h}\right)\left((\lambda_h^+ + \mu_h)S_h^+ - (\lambda_h+\mu_h)S_h\right) \\ &+ \left(1-\frac{A_h^+}{A_h}\right)\left(\lambda_h S_h - \frac{\lambda_h^+ S_h^+}{A_h^+}A_h\right) + c_1\left(1-\frac{I_h^+}{I_h}\right)\left(\gamma_h A_h - \frac{\gamma_h A_h^+}{I_h^+}I_h\right) \\ &+ c_2\left(1-\frac{S_r^+}{S_r}\right)\left((\lambda_r^+ + \mu_r)S_r^+ - (\lambda_r+\mu_r)S_r\right) + c_3\left(\lambda_r S_r - m_3 I_r\right),\end{aligned}$$

which implies

$$\begin{aligned}\frac{\partial L}{\partial t} =& \mu_h S_h^+\left(2-\frac{S_h}{S_h^+}-\frac{S_h^+}{S_h}\right) + \frac{\beta_1 A_h^+ S_h^+}{N_h^+}\left(2-\frac{S_h^+}{S_h}-\frac{A_h}{A_h^+}\right) \\ &+ \frac{\beta_1 A_h^+ S_h^+}{N_h}\left(\frac{A_h}{A_h^+}-\frac{S_h}{S_h^+}\right) + \frac{\beta_2 I_h^+ S_h^+}{N_h^+}\left(2-\frac{S_h^+}{S_h}-\frac{A_h}{A_h^+}\right) \\ &+ \frac{\beta_2 I_h^+ S_h^+}{N_h}\left(\frac{I_h}{I_h^+}-\frac{S_h A_h^+ I_h}{S_h^+ A_h I_h^+}\right) + \frac{\beta_2 I_h^+ S_h^+}{N_h^+}\left(1-\frac{I_h}{I_h^+}-\frac{A_h I_h^+}{A_h^+ I_h}+\frac{A_h}{A_h^+}\right) \\ &+ \frac{\beta_3 I_r S_h^+}{N_h}\left(1-\frac{S_h A_h^+}{S_h^+ A_h}\right) + \frac{\mu_r {S_r^+}^2}{\beta_4}\left(2-\frac{S_r}{S_r^+}-\frac{S_r^+}{S_r}\right) \\ &+ \frac{I_r {S_r^+}^2}{N_r}\left(1-\frac{S_r}{S_r^+}\right) + \frac{I_r S_r^+}{m_3-\beta_4}\left(\frac{\beta_4 S_r}{N_r}-m_3\right).\end{aligned}$$

By applying algebraic simplification under the assumptions $N_h = N_h^+$, $N_r = N_r^+$, $S_r \le S_r^+$, and inequalities such as $\frac{S_h^+}{S_h} + \frac{A_h}{A_h^+} \le 2$, we obtain

$$\begin{aligned}\frac{\partial L}{\partial t} \le& \mu_h S_h^+\left(2-\frac{S_h}{S_h^+}-\frac{S_h^+}{S_h}\right) + \frac{\beta_1 A_h^+ S_h^+}{N_h^+}\left(2-\frac{S_h}{S_h^+}-\frac{S_h^+}{S_h}\right) \\ &+ \frac{\beta_2 I_h^+ S_h^+}{N_h^+}\left(3-\frac{S_h^+}{S_h}-\frac{A_h I_h^+}{A_h^+ I_h}-\frac{S_h A_h^+ I_h}{S_h^+ A_h I_h^+}\right) \\ &+ \frac{\beta_3 I_r S_h^+}{N_h^+}\left(3-\frac{S_h^+}{S_h}-\frac{A_h}{A_h^+}-\frac{S_h A_h^+}{S_h^+ A_h}\right)\end{aligned}$$

$$+\frac{\mu_r {S_r^+}^2}{\beta_4}\left(2-\frac{S_r}{S_r^+}-\frac{S_r^+}{S_r}\right)-I_rS_r. \tag{24}$$

By utilizing the arithmetic-geometric mean inequality, we get

$$\left(2-\frac{S_h}{S_h^+}-\frac{S_h^+}{S_h}\right)\le 0, \qquad \left(3-\frac{S_h^+}{S_h}-\frac{A_hI_h^+}{A_h^+I_h}-\frac{S_hA_h^+I_h}{S_h^+A_hI_h^+}\right)\le 0,$$
$$\left(3-\frac{S_h^+}{S_h}-\frac{A_h}{A_h^+}-\frac{S_hA_h^+}{S_h^+A_h}\right)\le 0, \qquad \left(2-\frac{S_r}{S_r^+}-\frac{S_r^+}{S_r}\right)\le 0.$$

Therefore, $\frac{\partial L}{\partial t}<0$ if and only if $\mathcal{R}_0^r<1$, and $\frac{\partial L}{\partial t}=0$ only at the HEE. By applying LaSalle's invariance principle [21], we conclude that the HEE is globally asymptotically stable if and only if $\mathcal{R}_0^r<1$. □

Next, we analyze the global asymptotic stability of the EE, as stated in Theorem 5.

Theorem 5. Let the EE exist, which occurs when $\mathcal{R}_0^r>1$. The EE of model (1) is globally asymptotically stable.

Proof. Consider the Lyapunov function $L\colon D\to\mathbb{R}$ defined by

$$\begin{aligned}L=&\left(S_h-S_h^*-S_h^*\ln\frac{S_h}{S_h^*}\right)+\left(A_h-A_h^*-A_h^*\ln\frac{A_h}{A_h^*}\right)\\&+c\left(I_h-I_h^*-I_h^*\ln\frac{I_h}{I_h^*}\right)\\&+\left(S_r-S_r^*-S_r^*\ln\frac{S_r}{S_r^*}\right)+\left(I_r-I_r^*-I_r^*\ln\frac{I_r}{I_r^*}\right),\end{aligned} \tag{25}$$

where $c=\frac{\beta_2 I_h^*S_h^*}{\gamma_h A_h^*N_h^*}$.
Clearly, $L(X^*)=L_1^*+L_2^*+cL_3^*+L_4^*+L_5^*=0$, and $L>0$ for all $X\neq X^*$, as each component is positive definite. Differentiating L with respect to time yields

$$\begin{aligned}\frac{\partial L}{\partial t}=&\left(1-\frac{S_h^*}{S_h}\right)\frac{dS_h}{dt}+\left(1-\frac{A_h^*}{A_h}\right)\frac{dA_h}{dt}+c\left(1-\frac{I_h^*}{I_h}\right)\frac{dI_h}{dt}\\&+\left(1-\frac{S_r^*}{S_r}\right)\frac{dS_r}{dt}+\left(1-\frac{I_r^*}{I_r}\right)\frac{dI_r}{dt}.\end{aligned} \tag{26}$$

By using the same reasoning adopted in the proof of Theorem 4, then assuming $N_h=N_h^*$, $N_r=N_r^*$, and inequality $\frac{I_r}{I_r^*}+\frac{I_r^*}{I_r}\le 2$, we obtain

$$\frac{\partial L}{\partial t}\le \mu_hS_h^*\left(2-\frac{S_h}{S_h^*}-\frac{S_h^*}{S_h}\right)+\frac{\beta_1A_h^*S_h^*}{N_h^*}\left(2-\frac{S_h}{S_h^*}-\frac{S_h^*}{S_h}\right)$$

$$+\frac{\beta_2 I_h^* S_h^*}{N_h^*}\left(3-\frac{S_h^*}{S_h}-\frac{A_h I_h^*}{A_h^* I_h}-\frac{S_h A_h^* I_h}{S_h^* A_h I_h^*}\right)$$
$$+\frac{\beta_3 I_r^* S_h^*}{N_h^*}\left(4-\frac{S_h^*}{S_h}-\frac{A_h}{A_h^*}-\frac{I_r^*}{I_r}-\frac{S_h A_h^* I_r}{S_h^* A_h I_r^*}\right)$$
$$+\frac{\beta_4 I_r^* S_r^*}{N_r^*}\left(2-\frac{S_r}{S_r^*}-\frac{S_r^*}{S_r}\right)+\mu_r S_r^*\left(2-\frac{S_r}{S_r^*}-\frac{S_r^*}{S_r}\right). \tag{27}$$

By utilizing the arithmetic-geometric mean inequality, we get

$$\left(2-\frac{S_h}{S_h^*}-\frac{S_h^*}{S_h}\right)\le 0, \qquad \left(3-\frac{S_h^*}{S_h}-\frac{A_h I_h^*}{A_h^* I_h}-\frac{S_h A_h^* I_h}{S_h^* A_h I_h^*}\right)\le 0,$$
$$\left(4-\frac{S_h^*}{S_h}-\frac{A_h}{A_h^*}-\frac{I_r^*}{I_r}-\frac{S_h A_h^* I_r}{S_h^* A_h I_r^*}\right)\le 0, \qquad \left(2-\frac{S_r}{S_r^*}-\frac{S_r^*}{S_r}\right)\le 0.$$

Therefore, $\frac{\partial L}{\partial t}<0$ for all $X\neq X^*$, and $\frac{\partial L}{\partial t}=0$ only at the EE. By LaSalle's invariance principle [21], we conclude that the EE is globally asymptotically stable. □

4 Application of the model to real data

4.1 Parameter estimation

In this study, we utilized cumulative case data of Mpox infections in the United States (U.S.), spanning from June 1, 2022, to September 30, 2022, obtained from the official Centers for Disease Controls and Prevention (CDC) website [13]. According to the National Vital Statistics System, life expectancy in the United States in 2022 was 77.5 years, while the U.S. Census Bureau reported the total population as 333,287,557 individuals. Consequently, parameters π_h and μ_h are derived as follows:

$$\mu_h=\frac{1}{\text{life expectancy}}=\frac{1}{77.5\times 365}=\frac{1}{28287.5}=3.535\times 10^{-5}\ /\text{day},$$

$$\pi_h=\mu_h\times\text{total population}=\frac{333287557}{28287.5}=11782.149\ \text{individuals/day}.$$

The remaining parameters are estimated using the nonlinear least squares method. This procedure aimed to determine the optimal parameters by minimizing the sum of squared deviations between the observed data and the model predictions. The initial conditions are set as follows: $S_h(0)=333287548,\ A_h(0)=0,\ I_h(0)=9,\ R_h(0)=0,\ S_r(0)=800,\ I_r(0)=10.$

Figure 2 compares the cumulative number of Mpox infections from the reported data and the values from the fitted model. The model fits the empirical data well during most of the observation period. This agreement shows that the applied parameter set and the estimation strategy have captured the essence of the transmission dynamics of Mpox. Furthermore, the close proximity of the simulated curve to the reported data points reflects the robustness of the model in mimicking the epidemic curve of cumulative incidence, hence validating the suitability of the model for further forecasting and policy-oriented analyses.

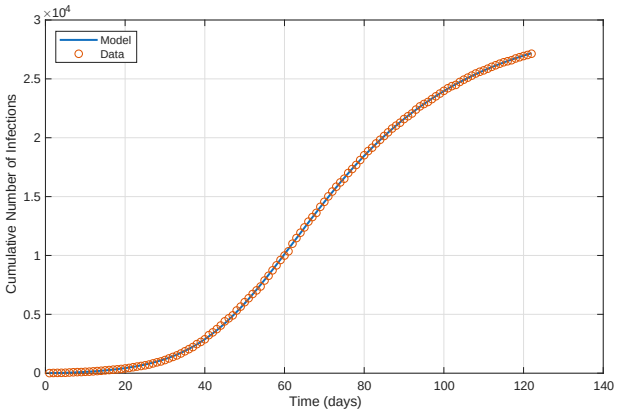


Figure 2: Comparison of model simulation and reported data for cumulative Mpox infections over time

The obtained mean absolute percentage error is 3.645%, indicating that the parameter estimation results exhibit a high level of accuracy and reliability. Referring to the estimated parameters presented in Table 3, the basic reproduction numbers are calculated as $\mathcal{R}_0^h = 0.88235$ for the human population and $\mathcal{R}_0^r = 2.08405$ for the rodent population, resulting in an overall basic reproduction number of $\mathcal{R}_0 = 2.08405$. Since the value of $\mathcal{R}_0$ exceeds unity, the model suggests that Mpox continues to persist in an endemic state within the United States.

Table 3: Parameter values

Parameter	Value	Parameter	Value
π_r	150.78	τ	0.11385
β_1	0.37198	δ_1	0.08249
β_2	0.07489	δ_2	0.07999
β_3	0.93455	δ_3	0.00011
β_4	0.19412	μ_r	0.09304
γ_h	0.60306		

4.2 Sensitivity analysis

In this section, a sensitivity analysis is carried out to determine the parameters that exert the greatest influence on the stability conditions of the equilibrium points in model (1). The sensitivity index corresponding to each parameter is evaluated using the relation $e_x = \frac{x}{\mathcal{R}_0} \times \frac{\partial \mathcal{R}_0}{\partial x}$, where x, $\mathcal{R}_0$, and e_x represent the parameter under consideration, the basic reproduction number, and the sensitivity index of x with respect to $\mathcal{R}_0$, respectively. A positive sensitivity index indicates that an increase (decrease) in the parameter value will lead to an increase (decrease) in $\mathcal{R}_0$. In contrast, a negative sensitivity index suggests that an increase (decrease) in the parameter value results in a decrease (increase) in $\mathcal{R}_0$. According to the parameter values in Table 3, the computed sensitivity indices for the parameters contributing to $\mathcal{R}_0$ are presented in Table 4.

Table 4: Sensitivity indices

Parameter	Sensitivity index on $\mathcal{R}_0^h$	Sensitivity index on $\mathcal{R}_0^r$
β_1	0.61491	–
β_2	0.38509	–
β_4	–	1
γ_h	−0.49454	–
τ	−0.22613	–
δ_1	−0.12032	–
δ_2	−0.15888	–
δ_3	–	−0.00118
μ_h	−0.00012	–
μ_r	–	−0.99882

As presented in Table 4, parameters β_1 and β_2 demonstrate positive sensitivity indices with respect to $\mathcal{R}_0^h$, implying that an increase in these parameters leads to a higher value of $\mathcal{R}_0^h$. Conversely, parameters γ_h, τ, δ_1, δ_2, and μ_h exhibit negative sensitivity indices, indicating that their increase reduces the value of $\mathcal{R}_0^h$. With regard to $\mathcal{R}_0^r$, only β_4 shows a positive sensitivity index, while δ_3 and μ_r display negative values. In general, β_1 and γ_h emerge as the most influential parameters on $\mathcal{R}_0^h$, as they possess the largest absolute sensitivity indices. The influence of parameters β_1 and γ_h is illustrated in Figure 3. For $\mathcal{R}_0^r$, the parameters with the greatest impact are β_4 and μ_r. Nevertheless, since the natural mortality rate of rodents (μ_r) is typically beyond human control, the most relevant intervention strategies should focus on β_4 and δ_3 within the rodent population.

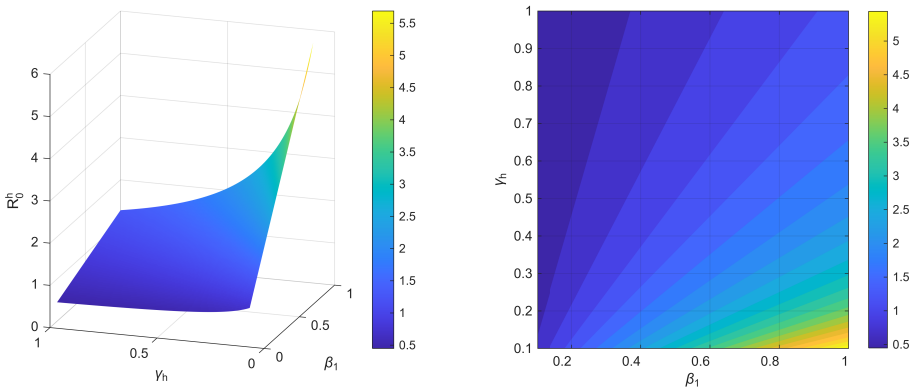


Figure 3: Influence of parameters β_1 and γ_h with respect to $\mathcal{R}_0^h$

4.3 Numerical simulation

In this section, numerical simulations are conducted to investigate the influence of parameters β_1 and γ_h on the dynamics of both A_h and I_h within the time horizon from $t = 0$ to $t = 500$ days. The outcomes of these simulations are illustrated in Figures 4 and 5. Figure 4 illustrates the temporal dynamics of the asymptomatic (A_h) and symptomatic (I_h) human subpopulations under varying values of the transmission parameter β_1. As observed, higher values of β_1 substantially accelerate the epidemic growth rate and lead to significantly larger equilibrium levels in both compartments. Specifically, when $\beta_1 = 0.4$, the prevalence of A_h and I_h reaches markedly higher peaks compared to scenarios with $\beta_1 = 0.3$ and $\beta_1 = 0.35$, followed by a gradual decline over time. On the other hand, smaller values of β_1 weaken the speed of an epidemic and stabilize with lower intensities, indicating that transmission intensity significantly controls the magnitude and persistence of infection in humans. These observations highlight that β_1 plays a paramount role in shaping disease burden and long-term epidemic outcomes.

The short-term dynamics (right panels) of Figure 5 illustrate the early behavior of the asymptomatic (A_h) and symptomatic (I_h) human subpopulations under varying values of the transition rate γ_h, which represents the progression rate of Mpox from the asymptomatic to symptomatic phase. For A_h, lower γ_h values (e.g., 0.1 and 0.3) lead to a slower depletion of the asymptomatic compartment, allowing a higher accumulation of infectious individuals and a sustained transmission potential. In contrast, higher γ_h (e.g., 1) rapidly converts A_h into I_h, leading to an initially faster rise in symptomatic cases. However, for I_h, this trend reverses over time: Although high γ_h causes an early peak in symptomatic individuals, the rapid depletion of A_h reduces the continued inflow into I_h, resulting in an eventual decline or stabilization. Meanwhile, lower γ_h values

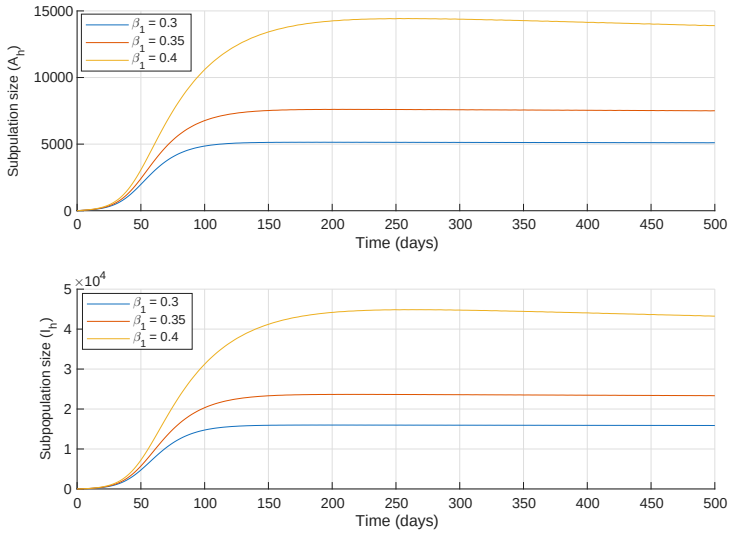


Figure 4: Influence of parameter β_1 with respect to A_h and I_h

prolong the infectious duration of asymptomatic carriers, thereby sustaining transmission and gradually increasing the I_h population to surpass the high γ_h scenarios. The long-term dynamics (left panels) confirm this reversal: Lower γ_h values sustain elevated equilibrium levels of both A_h and I_h, while higher γ_h values result in early saturation and lower prevalence. These results emphasize a nonlinear trade-off in γ_h: Higher values accelerate clinical progression but suppress long-term transmission, while lower values prolong asymptomatic infectious periods, yielding a higher cumulative disease burden.

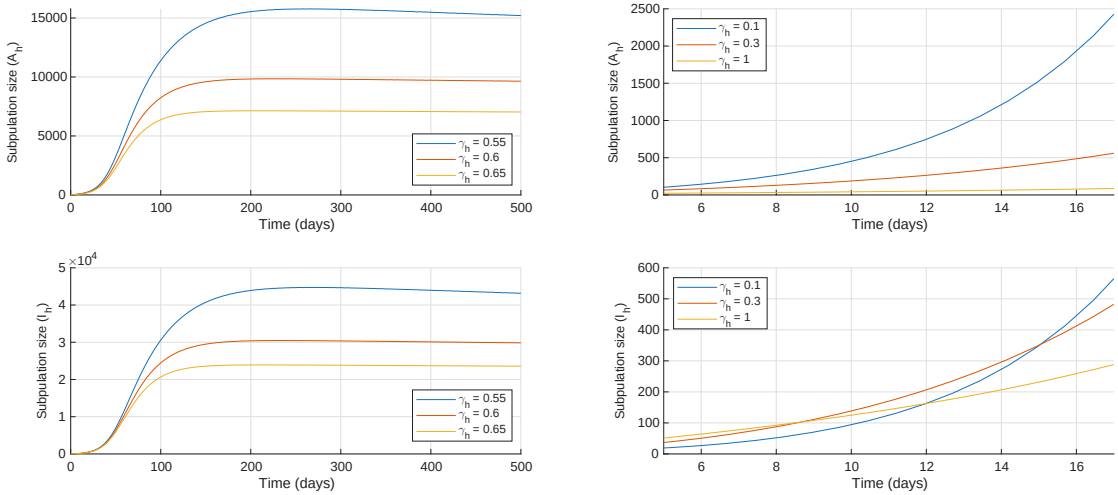


Figure 5: Influence of parameter γ_h with respect to A_h and I_h

Figure 6 depicts the temporal dynamics of the different human (S_h, A_h, I_h, R_h) and rodent (S_r, I_r) compartments from the simulated Mpox transmission model. Among the human subpopulations, the susceptible class S_h exhibits a gradual decrease in the temporal domain, whereas the recovered class R_h displays a steady increase, indicating the build-up of immunity over time. The asymptomatic (A_h) and symptomatic (I_h) compartments experience an initial increase, where I_h rises faster and remains consistently at higher levels than A_h, indicating that the majority of the human infectious pool comprises symptomatic infections. Meanwhile, the rodent dynamics exhibit a typical transient profile. The susceptible rodent pool S_r first increases, presumably due to births mediated by natural processes, then quickly decreases as infection occurs among the rodent population, and finally settles into a steady-state solution, indicating an endemic state. The infected rodent pool I_r increases during the initial phase of the outbreak, reflecting a rapid infection growth rate, then tapers off, further indicating that infection is persisting within the rodent reservoir. This suggests that interspecies interactions drive the spread of Mpox and points out the necessity of monitoring both host populations for long-term epidemic control.

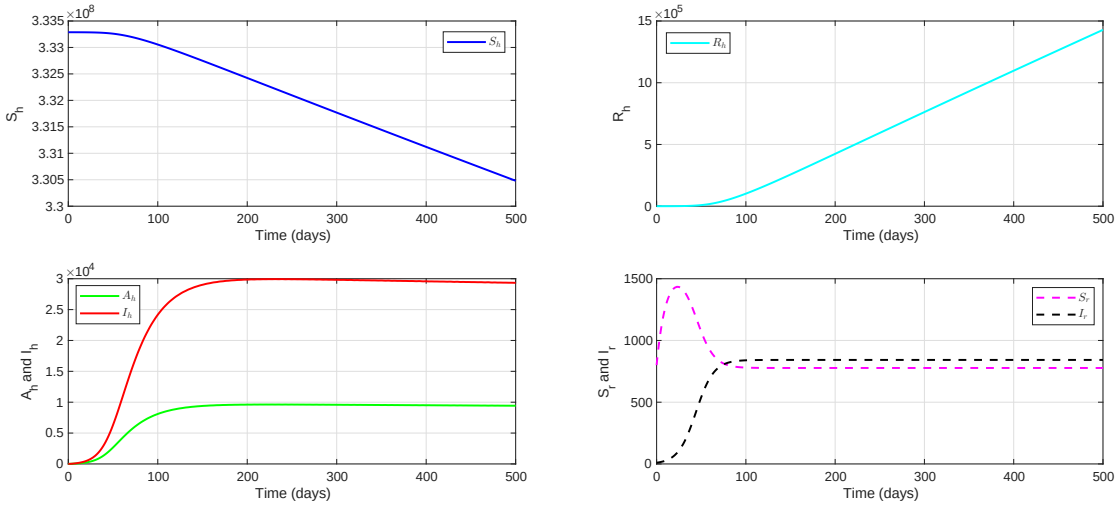


Figure 6: Dynamics of human and rodent populations

5 Optimal control

5.1 Model and characterization

This section investigates the incorporation of control measures into model (1) and subsequently analyzes the characterization of the corresponding optimal control problem. Two control measures are considered, namely vaccination and rodenticide usage. The vaccination control is represented by the time-dependent control variable $u_1(t)$, which describes vaccine administration to susceptible individuals at rate κ_1. Vaccines such as JYNNEOS are capable of inducing immunity after a certain period, thereby reducing the likelihood of infection among susceptible individuals [5]. The rodenticide usage control is represented by the time-dependent control variable $u_2(t)$, which corresponds to the targeted removal of Mpox-infected rodents at rate κ_2. This control has been recommended as a preventive strategy to limit Mpox transmission, as reported by the Leicester Mercury newspaper [23]. By incorporating these two controls, model (1) is reformulated as follows:

$$\begin{aligned}
\frac{dS_h}{dt} &= \pi_h - (\lambda_h + \mu_h + \kappa_1 u_1(t))S_h,\\
\frac{dA_h}{dt} &= \lambda_h S_h - (\gamma_h + \delta_1 + \mu_h)A_h,\\
\frac{dI_h}{dt} &= \gamma_h A_h - (\tau + \delta_2 + \mu_h)I_h,\\
\frac{dR_h}{dt} &= \kappa_1 u_1(t) S_h + \tau I_h - \mu_h R_h,\\
\frac{dS_r}{dt} &= \pi_r - (\lambda_r + \mu_r)S_r,\\
\frac{dI_r}{dt} &= \lambda_r S_r - (\delta_3 + \mu_r + \kappa_2 u_2(t))I_r,
\end{aligned} \tag{28}$$

where

$$\lambda_h = \frac{\beta_1 A_h + \beta_2 I_h + \beta_3 I_r}{N_h}, \qquad \lambda_r = \frac{\beta_4 I_r}{N_r}.$$

The optimal control problem associated with model (28) is formulated to reduce the number of asymptomatic humans (A_h), symptomatic humans (I_h), and infected rodents (I_r), while simultaneously minimizing the implementation costs of the applied controls. The vaccination control $u_1(t)$ generates costs related to vaccine procurement, distribution, and healthcare personnel. The rodenticide usage control $u_2(t)$ incurs operational costs related to labor, equipment, and chemical procurement expenses. Accordingly, the objective functional is defined by

$$F(\mathbf{x}(t), \mathbf{u}(t)) = \int_0^{t_f} \left(v_1 A_h + v_2 I_h + v_3 I_r + \frac{1}{2} \left(w_1 u_1^2(t) + w_2 u_2^2(t) \right) \right) dt, \tag{29}$$

where v_1, v_2, v_3 denote the weighting coefficients associated with the infected compartments, whereas w_1, w_2 denote the cost coefficients corresponding to the implemented control measures. The quadratic expressions $u_1^2(t)$ and $u_2^2(t)$ are introduced to preserve the convexity of the objective functional, thereby facilitating analytical treatment and ensuring the existence of an optimal solution. In addition, this formulation captures the realistic assumption that higher control intensities produce disproportionately greater implementation costs. Such a quadratic structure is also commonly employed in optimal control theory because it simplifies the derivation of optimality conditions through Pontryagin's principle. The objective is therefore to identify the optimal controls $u_1^*(t)$ and $u_2^*(t)$ that minimize the objective functional, namely,

$$F(\mathbf{x}(t), \mathbf{u}^*(t)) = \min \left\{ F(\mathbf{x}(t), \mathbf{u}(t)) \mid (u_1(t), u_2(t)) \in U \right\},$$

where the admissible control set is defined as

$$U = \left\{ (u_1(t), u_2(t)) \mid 0 \leq u_1(t) \leq 1,\ 0 \leq u_2(t) \leq 1 \right\}.$$

The optimal control problem corresponding to model (28) is established using Pontryagin's principle [14]. By employing this principle, the objective functional in (29) is transformed into the Hamiltonian function $\mathcal{H}$, which is defined as follows:

$$\begin{aligned} \mathcal{H}(\mathbf{x}(t), \mathbf{u}(t), \mathbf{p}(t), t) =& v_1 A_h + v_2 I_h + v_3 I_r + \frac{1}{2} \left(w_1 u_1^2(t) + w_2 u_2^2(t) \right) \\ &+ p_1(t)(\pi_h - (\lambda_h + \mu_h + \kappa_1 u_1(t)) S_h) \\ &+ p_2(t)(\lambda_h S_h - (\gamma_h + \delta_1 + \mu_h) A_h) \\ &+ p_3(t)(\gamma_h A_h - (\tau + \delta_2 + \mu_h) I_h) \\ &+ p_4(t)(\kappa_1 u_1(t) S_h + \tau I_h - \mu_h R_h) \\ &+ p_5(t)(\pi_r - (\lambda_r + \mu_r) S_r) \\ &+ p_6(t)(\lambda_r S_r - (\delta_3 + \mu_r + \kappa_2 u_2(t)) I_r), \end{aligned} \tag{30}$$

where $p_i(t)$ for $i = 1, 2, \ldots, 6$ denote the costate variables.

Based on Pontryagin's principle, an optimal solution satisfies the conditions given by the state equations, costate equations, and stationary conditions of the Hamiltonian function. The state equations in model (28) are obtained from the partial derivatives of $\mathcal{H}$ with respect to the costate variables. Meanwhile, the costate equations are obtained from the negative partial derivatives of

$\mathcal{H}$ with respect to the state variables and are formulated as

$$\begin{aligned}
\frac{dp_1(t)}{dt} &= (p_1(t)-p_2(t))\left(\lambda_h - \frac{\lambda_h S_h}{N_h}\right) + p_1(t)(\mu_h + \kappa_1 u_1(t)) - p_4(t)\kappa_1 u_1(t),\\
\frac{dp_2(t)}{dt} &= -v_1 + (p_1(t)-p_2(t))\left(\frac{\beta_1 - \lambda_h}{N_h}\right) S_h + p_2(t)(\gamma_h + \delta_1 + \mu_h)\\
&\quad - p_3(t)\gamma_h,\\
\frac{dp_3(t)}{dt} &= -v_2 + (p_1(t)-p_2(t))\left(\frac{\beta_2 - \lambda_h}{N_h}\right) S_h + p_3(t)(\tau + \delta_2 + \mu_h) - p_4(t)\tau,\\
\frac{dp_4(t)}{dt} &= (p_2(t)-p_1(t))\frac{\lambda_h S_h}{N_h} + p_4(t)\mu_h,\\
\frac{dp_5(t)}{dt} &= (p_5(t)-p_6(t))\left(\lambda_r - \frac{\lambda_r S_r}{N_r}\right) + p_5(t)\mu_r,\\
\frac{dp_6(t)}{dt} &= -v_3 + (p_1(t)-p_2(t))\frac{\beta_3 S_h}{N_h} + (p_5(t)-p_6(t))\left(\frac{\beta_4 - \lambda_r}{N_r}\right) S_r\\
&\quad + p_6(t)(\delta_3 + \mu_r + \kappa_2 u_2(t)),
\end{aligned} \tag{31}$$

subject to the transversality conditions $p_i(t_f) = 0$ for $i = 1, 2, \ldots, 6$. These conditions follow from Pontryagin's principle because t_f is fixed, $\mathbf{x}(t_f)$ is free, and the objective functional contains no terminal payoff term. Consequently, the costate variables p_i become zero at t_f, indicating that the marginal contribution of each state variable to the objective functional becomes zero at the end of the control period.

In addition, the stationary conditions are obtained by differentiating $\mathcal{H}$ with respect to the control variables. Therefore, the control functions $u_1(t)$ and $u_2(t)$ are represented by

$$u_1(t) = \frac{(p_1(t)-p_4(t))\kappa_1 S_h}{w_1}, \qquad u_2(t) = \frac{p_6(t)\kappa_2 I_r}{w_2}.$$

To maintain biological admissibility, each control variable is constrained within the interval $[0, 1]$. Therefore, the optimal control profiles $u_1^*(t)$ and $u_2^*(t)$ are determined as

$$\begin{aligned}
u_1^*(t) &= \min\left\{1, \max\left\{0, \frac{(p_1(t)-p_4(t))\kappa_1 S_h}{w_1}\right\}\right\},\\
u_2^*(t) &= \min\left\{1, \max\left\{0, \frac{p_6(t)\kappa_2 I_r}{w_2}\right\}\right\}.
\end{aligned} \tag{32}$$

5.2 Simulation with control

The optimal control problem associated with model (28) is solved numerically by employing the forward-backward sweep method. The parameter values presented in Table 3 are adopted in the simulations within the time horizon from $t = 0$ to $t = 500$ days. To investigate the impact of different intervention policies, three control strategies are considered in the numerical simulations. These strategies are formulated based on two control variables, namely vaccination ($u_1(t)$) and rodenticide usage ($u_2(t)$). Strategy A considers the implementation of vaccination only ($u_1(t)$), Strategy B applies rodenticide usage only ($u_2(t)$), whereas Strategy C combines vaccination and rodenticide usage simultaneously ($u_1(t), u_2(t)$). Each control strategy is evaluated by comparing it with the uncontrolled scenario to examine its effectiveness in reducing the asymptomatic human, symptomatic human, and infected rodent subpopulations.

Figure 7 illustrates the dynamics of the asymptomatic human subpopulation (A_h) under all control strategies compared with the uncontrolled scenario. It can be observed that, in the absence of control strategies, the value of A_h increases sharply from the beginning of the simulation period until approximately day 100, after which the growth rate gradually slows and eventually stabilizes until the end of the simulation period. This behavior indicates a rapid increase in A_h, thereby elevating the risk of Mpox transmission. In contrast, the implementation of control strategies significantly reduces the number of A_h, demonstrating the effectiveness of control strategies in suppressing the growth of the asymptomatic human subpopulation. Among the considered strategies, Strategy C, which combines both control strategies, appears to be the most effective, although its difference from Strategy A is relatively small.

A similar pattern is observed in the dynamics of the symptomatic human subpopulation (I_h), as illustrated in Figure 8. The value of I_h also increases rapidly in the uncontrolled scenario, while Strategy C remains the most effective approach for suppressing the growth of I_h. Furthermore, Figure 9, which presents the dynamics of the infected rodent subpopulation (I_r), exhibits behavior distinct from that of A_h and I_h. The application of Strategy A has no effect on the dynamics of I_r, as its trajectory coincides with the uncontrolled scenario. This occurs because vaccination is administered only to the susceptible human subpopulation, and there is no transmission pathway from humans to rodents. Meanwhile, Strategies B and C are highly effective in reducing the growth of I_r.

The control profiles for all strategies are presented in Figure 10. Under Strategy A, the control variable $u_1(t)$ remains at its upper bound throughout the entire simulation period. In contrast, under Strategy B, the control variable $u_2(t)$ decreases sharply starting around day 25 and reaches zero from approximately day 35 until the end of the simulation period. Under

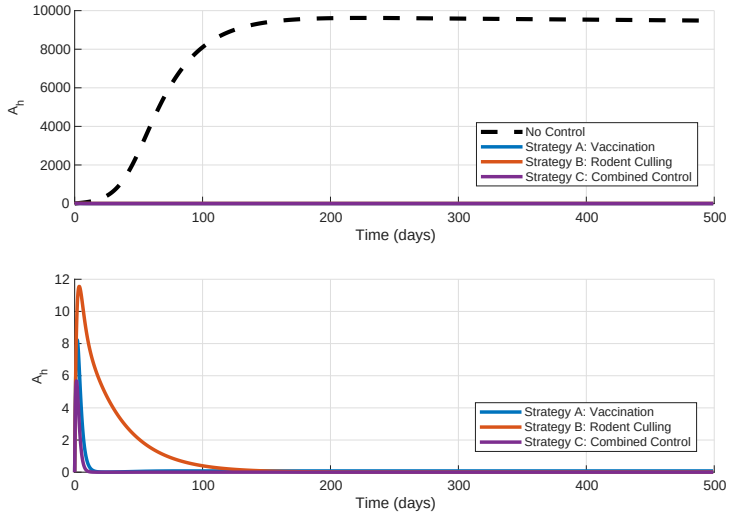


Figure 7: Dynamics of A_h under all control strategies

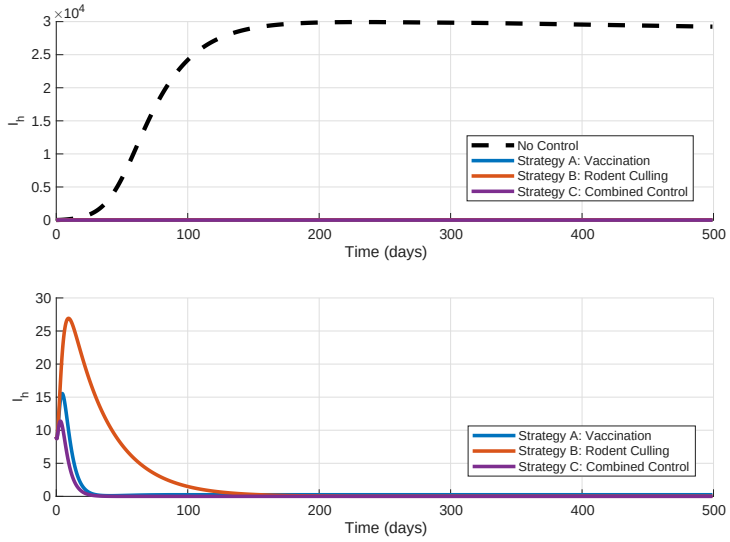


Figure 8: Dynamics of I_h under all control strategies

Strategy C, the control variable $u_1(t)$ decreases sharply first, followed by $u_2(t)$, and both controls reach zero before day 50 and remain at zero until the end of the simulation period.

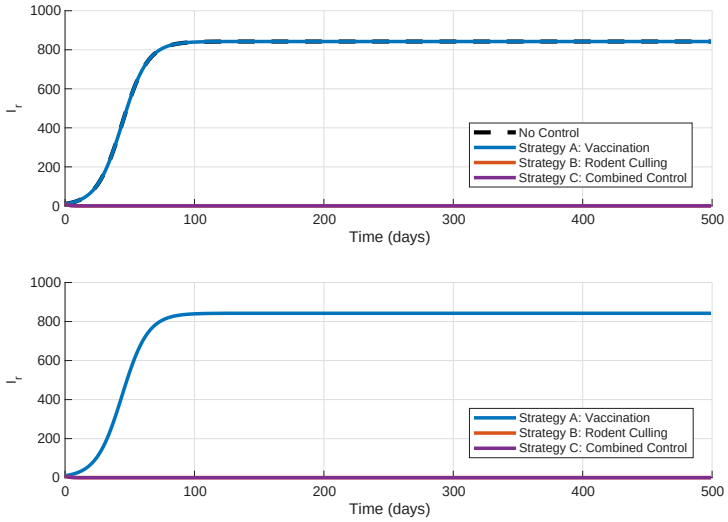


Figure 9: Dynamics of I_r under all control strategies

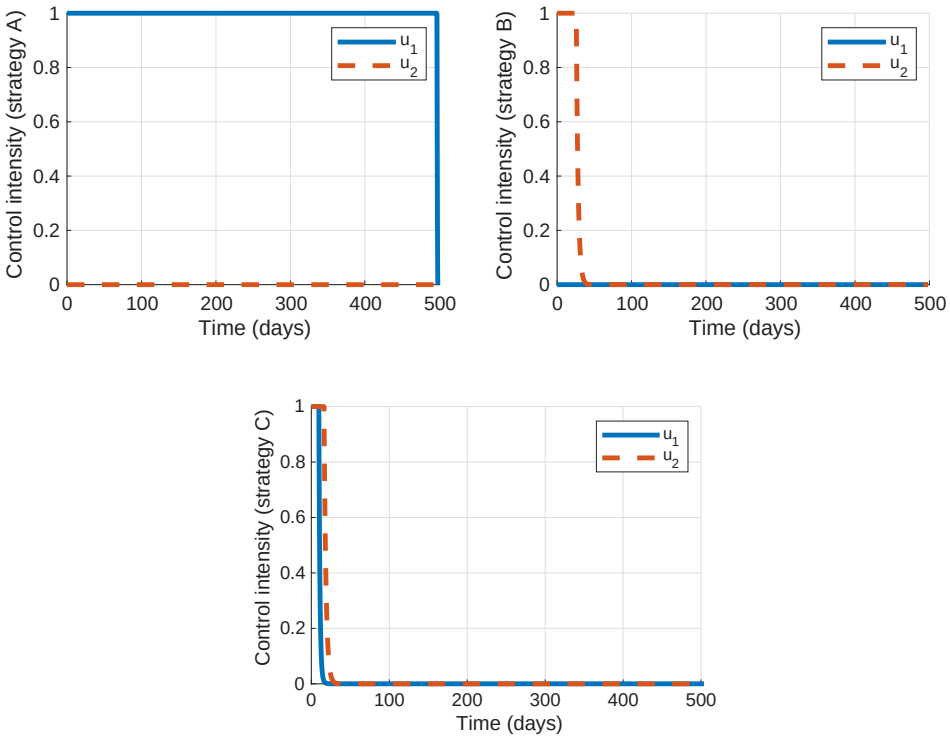


Figure 10: Control profiles for all strategies

5.3 Cost-effectiveness

Cost-effectiveness analysis evaluates the economic efficiency of control strategies by comparing their implementation costs with the corresponding health outcomes. In this section, we employ the incremental cost-effectiveness ratio (ICER) to identify the strategy that provides the greatest health benefit relative to its cost. Following [19], the ICER is defined as

$$\text{ICER} = \frac{\text{Difference in total cost between strategies } i \text{ and } j}{\text{Difference in total infections averted between strategies } i \text{ and } j}.$$

The total infections averted are estimated by subtracting the total infections occurring after the implementation of a specific control strategy from the projected total infections in the absence of any strategy. The values of the total cost and total infections averted for all strategies are presented in Table 5.

Table 5: Total cost and total infections averted

Strategy	Total cost	Total infections averted
A	13,657	1,222,500
B	24,701	1,268,800
C	12,494	1,269,500

The control strategies are first ranked according to the total number of infections averted in ascending order, starting from the strategy with the lowest number of infections averted. Subsequently, the ICER values can be calculated sequentially in the following manner:

$$\begin{aligned}
\text{ICER (A)} &= \frac{13{,}657 - 0}{1{,}222{,}500 - 0} = 0.0112,\\
\text{ICER (B)} &= \frac{24{,}701 - 13{,}657}{1{,}268{,}800 - 1{,}222{,}500} = 0.2385,\\
\text{ICER (C)} &= \frac{12{,}494 - 24{,}701}{1{,}269{,}500 - 1{,}268{,}800} = -17.4386.
\end{aligned}$$

The initial ICER calculation is performed using the ranked strategies, where Strategy B is identified as having the highest ICER value. Therefore, Strategy B is excluded from the subsequent analysis, and the ICER values are recalculated using the same approach for the remaining strategies, which are then ordered as Strategies A and C. The next ICER calculation is presented as follows:

$$\text{ICER (A)} = \frac{13{,}657 - 0}{1{,}222{,}500 - 0} = 0.0112,$$
$$\text{ICER (C)} = \frac{12{,}494 - 13{,}657}{1{,}269{,}500 - 1{,}222{,}500} = -0.0247.$$

Based on this result, Strategy C has a lower ICER value than Strategy A, suggesting that Strategy C is the most cost-effective strategy.

6 Conclusion

The SAIR-SI model of Mpox transmission developed in this study comprises six compartments: Susceptible humans (S_h), asymptomatic humans (A_h), symptomatic infected humans (I_h), recovered humans (R_h), susceptible rodents (S_r), and infected rodents (I_r). The model satisfies the well-posedness, boundedness, and nonnegativity properties. Three equilibrium points are identified: The DFE, the HEE, and the EE. The DFE is both locally and globally asymptotically stable when the basic reproduction number $\mathcal{R}_0$ is less than one. The HEE is globally asymptotically stable provided that $\mathcal{R}_0^r$ is less than one. If the EE exists, then it is globally asymptotically stable. Sensitivity analysis reveals that parameters β_1 and γ_h exert the greatest influence on $\mathcal{R}_0^h$, with β_1 positively correlated and γ_h negatively correlated with $\mathcal{R}_0^h$. Conversely, parameters β_4 and δ_3 most significantly affect $\mathcal{R}_0^r$, where β_4 shows a positive correlation and δ_3 shows a negative correlation. The results of this study reveal that key transmission parameters drive the dynamics within both human and rodent populations during outbreaks. In detail, transmission parameter β_1 significantly influences the growth rate of epidemics and equilibrium levels; higher values accelerate epidemic progression, leading to larger peaks in asymptomatic and symptomatic subpopulations, A_h and I_h, respectively. Progression rate γ_h is equally crucial, as lower values prolong asymptomatic infectious periods, thus sustaining transmission and yielding a higher cumulative disease burden, while higher values expedite clinical progression but suppress long-term transmission. Furthermore, two control measures, namely vaccination and rodenticide usage, are introduced into the model. The characterization of the optimal control is analyzed using Pontryagin's principle. Three control strategies are considered in the numerical simulations: Strategy A involves vaccination only, Strategy B applies rodenticide usage only, and Strategy C combines both control measures simultaneously. The simulation results indicate that Strategy C is the most effective approach for reducing the asymptomatic human, symptomatic human, and infected rodent subpopulations. The cost-effectiveness analysis confirms that Strategy C is also the most efficient strategy, providing valuable guidance for policymakers

in selecting a control strategy that achieves improved health outcomes while considering limited available resources.

Further research will focus on addressing the limitations of the current model by incorporating additional epidemiological factors, including age structure, sexual transmission stratification, seasonal variations, and heterogeneous mixing patterns. Moreover, the inclusion of supportive treatment control and environmental influences may further improve the model framework and enhance its predictive capability for Mpox transmission dynamics.

Acknowledgments

The authors would like to express their sincere gratitude to all those who contributed to the success of this study, including the valuable feedback from the reviewers and the editorial support provided by the journal team.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Abidemi, A., Fatmawati and Peter, O.J. *An optimal control model for dengue dynamics with asymptomatic, isolation, and vigilant compartments*, Decis. Anal. J. 10 (2024), 100413.

[2] Alakunle, E., Moens, U., Nchinda, G. and Okeke, M.I. *Monkeypox virus in Nigeria: infection biology, epidemiology, and evolution*, Viruses 12 (2020), 1257.

[3] Alkhazzan, A., Wang, J., Nie, Y., Shah, S.M.A., Almutairi, D.K., Khan, H. and Alzabut, J. *Lyapunov-based analysis and worm extinction in wireless networks using stochastic SVEIR model*, Alex. Eng. J. 118 (2025), 337–353.

[4] Allehiany, F.M., Darassi, M.H., Ahmad, I., Khan, M.A. and Tag-eldin, E.M. *Mathematical modeling and backward bifurcation in monkeypox disease under real observed data*, Results Phys. 50 (2023), 106557.

[5] Alshahrani, A.M. *Cost-benefit analysis of interventions to mitigate the monkeypox virus*, Int. J. Environ. Res. Public Health 19 (2022), 13789.

[6] Alshehri, A.M. and Ullah, S. *Optimal control analysis of monkeypox disease with the impact of environmental transmission*, AIMS Math. 8 (2023), 16926–16960.

[7] Al-Shomrani, M.M., Musa, S.S. and Yusuf, A. *Unfolding the transmission dynamics of monkeypox virus: an epidemiological modelling analysis*, Mathematics 11 (2023), 1121.

[8] Bragazzi, N.L., Han, Q., Iyaniwura, S.A., Omame, A., Shausan, A., Wang, X., Woldegerima, W.A., Wu, J. and Kong, J.D. *Adaptive changes in sexual behavior in the high-risk population in response to human monkeypox transmission in Canada can help control the outbreak: insights from a two-group, two-route epidemic model*, J. Med. Virol. 95(4) (2023), e28575.

[9] Brauer, F., Castillo-Chavez, C. and Feng, Z. *Endemic disease models*, in: *Mathematical Models in Epidemiology*, Springer, 2019.

[10] CDC. *About Monkeypox*, 2026. Available at: https://www.cdc.gov/monkeypox/about/index.html

[11] CDC. *FACT SHEET: United States Response to the Clade I Mpox Outbreak in Several African Countries*, 2024. Available at: https://www.cdc.gov/media/releases/2024/s0822-mpox-outbreak.html

[12] CDC. *How Monkeypox Spreads*, 2026. Available at: https://www.cdc.gov/monkeypox/causes/index.html

[13] CDC. *U.S. Case Data: Monkeypox*, 2026. Available at: https://www.cdc.gov/monkeypox/data-research/cases/index.html

[14] Chen, C. and Xiang, Z. *Analysis and optimal control strategies of epidemic models with nonlinear incidence rates*, Adv. Contin. Discrete Models 2025 (2025), 92.

[15] Elmekawy, A., El-Azab, M.S. and El-Shenawy, A. *SIR dengue fever nonlinear mathematical model: a computational framework with numerical simulation*, Nonlinear Sci. 5 (2025), 100073.

[16] Elsonbaty, A., Adel, W., Aldurayhim, A. and El-Mesady, A. *Mathematical modeling and analysis of a novel monkeypox virus spread integrating imperfect vaccination and nonlinear incidence rates*, Ain Shams Eng. J. 15 (2024), 102451.

[17] Fernandes, G.D. and Maldonado, V. *Behavioral aspects and the transmission of monkeypox: a novel approach to determine the probability of transmission for sexually transmissible diseases*, Infect. Dis. Model. 8 (2023), 842–854.

[18] Flores-Pérez, A., González-Olvera, M.A., Chávez-Peña, G., Gallardo-Hernández, A.G. and Torres, L. *Time-delay enhanced SIR model for COVID-19 waves in Mexico: parameter estimation using evolutionary algorithms*, J. Theor. Biol. 613 (2025), 112229.

[19] Janan, T., Fatmawati, Alfiniyah, C., Martini, S., Aldila, D. and Hasan, A. *Dynamical analysis and optimal control of the Mpox transmission model with stratified susceptibility*, Math. Model. Numer. Simul. Appl. 6 (2026), 186–227.

[20] Janan, T., Fatmawati and Hasan, A. *Dynamical Analysis of Mpox Transmission Model Incorporating Asymptomatic Individuals*, Commun. Biomath. Sci. 9 (2026), 48–67.

[21] La Salle, J.P. *The Stability of Dynamical Systems*, SIAM, 1976.

[22] Leandry, L. and Mureithi, E.W. *An investigation on the monkeypox virus dynamics in human and rodent populations for a deterministic mathematical model*, Informatics Med. Unlocked 41 (2023), 101325.

[23] Leicester Mercury. *Rodent pets could be culled after UK monkeypox outbreak*, 2022. Available at: https://www.leicestermercury.co.uk/news/uk-world-news/rodent-pets-could-culled-after-7134597

[24] Li, S., Samreen, Ullah, S., AlQahtani, S.A., Tag, S.M. and Akgül, A. *Mathematical assessment of monkeypox with asymptomatic infection: prediction and optimal control analysis with real data application*, Results Phys. 51 (2023), 106726.

[25] Meetei, M.Z., DarAssi, M.H., Ahmad, I., Khan, M.A., Isa, N.M. and Alzahrani, E. *Mathematical modeling and optimal control analysis of the monkeypox Clade II infection dynamics under the recent cases of USA*, Comput. Ind. Eng. 209 (2025), 111462.

[26] Murray, J.D. *Mathematical Biology: I. An Introduction*, Springer, 2007.

[27] Naidu, D.S. *Optimal Control Systems*, Taylor & Francis, 2002.

[28] Somma, S.A., Akinwande, N.I. and Chado, U.D. *A mathematical model of monkeypox virus transmission dynamics*, Ife J. Sci. 21 (2019), 195–204.

[29] Stephano, M.A., Mayengo, M.M. and Irunde, J.I. *The role of asymptomatic carriers on the dynamics of a lymphatic filariasis model incorporating control strategies*, Results Control Optim. 15 (2024), 100425.

[30] Thornhill, J.P., Barkati, S., Walmsley, S., Rockstroh, J., Antinori, A., Harrison, L.B., Palich, R., Nori, A., Reeves, I. and Habibi, M.S. *Monkeypox virus infection in humans across 16 countries–April-June 2022*, N. Engl. J. Med. 387 (2022), 679–691.

[31] Usman, S. and Adamu, I.I. *Modeling the transmission dynamics of the monkeypox virus infection with treatment and vaccination interventions*, J. Appl. Math. Phys. 5 (2017), 2335–2353.

[32] US. *FACT SHEET: The United States Commitment to Address the Global Mpox Outbreak*, 2024. Available at: https://geneva.usmission.gov/2024/09/25/u-s-commitment-to-address-the-global-mpox-outbreak/

[33] Waseel, F., Streftaris, G., Rudrusamy, B. and Dass, S.C. *Assessing the dynamics and impact of COVID-19 vaccination on disease spread: a data-driven approach*, Infect. Dis. Model. 9(2) (2024), 527–556.

[34] WHO. *Mpox (monkeypox) outbreak 2022*, 2022. Available at: https://www.who.int/emergency/situations/monkeypox-oubreak-2022

[35] WHO. *Mpox*, 2024. Available at: https://www.who.int/news-room/fact-sheets/detail/mpox

[36] WHO. *Mpox*, 2024. Available at: https://www.who.int/news-room/questions-and-answers/item/mpox

[37] WHO. *WHO Director-General declares mpox outbreak a public health emergency of international concern*, 2024. Available at: https://www.who.int/news/item/14-08-2024-who-director-general-declares-mpox-outbreak-a-public-health-\emergency-of-international-concern

[38] Winata, I.G.S., Leonardo, L., Sylfiasari, R., Ekafentie, A., Immanuel, S.S. and Christyani, F. *Unveiling the outcomes of Mpox in pregnancy: a systematic review and single-arm meta-analysis*, Infect. Chemother. 57(1) (2025), 119–130.

[39] Wireko, F.A., Adu, I.K., Sebil, C. and Asamoah, J.K.K. *A fractal-fractional order model for exploring the dynamics of monkeypox disease*, Decis. Anal. J. 8 (2023), 100300.

[40] Xu, T., Zhang, J. and Jin, Z. *Analysis of final size and peak time for SIR epidemic model on simplicial complexes*, Chaos Solitons Fractals 196 (2025), 116390.

[41] Yapışkan, D., Yurtoğlu, M., Avcı, D., İskender Eroğlu, B.B. and Bonyah, E. *A novel model for monkeypox disease: system analysis and optimal preventive strategies*, Iran. J. Sci. 47 (2023), 1665–1677.

[42] Zhang, X.S., Mandal, S., Mohammed, H., Turner, C., Florence, I., Walker, J., Niyomsri, S., Amirthalingam, G., Ramsay, M., Charlett, A. and Vickerman, P. *Transmission dynamics and effect of control measures on the 2022 outbreak of Mpox among gay, bisexual, and other men who have sex with men in England: a mathematical modelling study*, Lancet Infect. Dis. 24(1) (2024), 65–74.

Iranian Journal of Numerical Analysis and Optimization
Vol. 17, No. 1, 2027, pp 416–456
https://doi.org/10.22067/ijnao.2026.97215.1817
https://ijnao.um.ac.ir/

CC BY NC

Research Article

Stability analyses of block hybrid methods for numerical solutions of differential equations

O. Adedire* and U.W. Sirisena

Abstract

This research presents a comparative analysis of stability polynomials and stability functions for investigating the stability of derived block hybrid methods. The methods—an explicit (BHEM) and two implicit (BHIM) formulations—were obtained through continuous formulation using multistep collocation and matrix inversion techniques. A novel theoretical result establishes that for consistent block methods, stability polynomials and stability functions must yield congruent stability conclusions, with any discrepancy indicating potential derivation errors. It is demonstrated that the use of stability polynomials offers a straightforward determination of the angle α for methods that are $A(\alpha)$-stable. However, for investigating stronger stability properties such as L-stability and $L(\alpha)$-stability, the stability function is preferable due to its direct rational form, which allows immediate asymptotic analysis. Detailed experiments on block methods and block hybrid methods reveal the respective advantages and limitations of each stability analysis tool. The analysis conclusively answers the research question: The absence of poles in the left half-plane alone does not guarantee A-stability; both analyticity

*Corresponding author

Received 27 December 2025; revised 30 June 2026; accepted 6 July 2026

Oludare Adedire
Department of Mathematics, Faculty of Natural Sciences, University of Jos, Plateau State, Nigeria; and Department of Statistics, Federal College of Forestry Jos. e-mail: dharenss@gmail.com

Ullusihewage Wipuni Sirisena
Department of Mathematics, Faculty of Natural Sciences, University of Jos, Plateau State, Nigeria. e-mail: swipuni@yahoo.com

How to cite this article
Adedire, O. and Sirisena, U.W., Stability analyses of block hybrid methods for numerical solutions of differential equations. *Iran. J. Numer. Anal. Optim.*, 2027; 17(1): 416-456. https://doi.org/10.22067/ijnao.2026.97215.1817

in the left half-plane and boundedness on the imaginary axis are required conditions. This research successfully extends stability analysis from the one-step domain to multi-step block formulations.

AMS subject classifications (2020): Primary 65C20; Secondary 65L07, 65L20, 65Y20.

Keywords: Block methods; Absolute stability regions; Congruence for consistent block methods; Stability polynomial; Stability function.

1 Introduction

The literature contains numerous numerical methods developed for solving differential equations (DEs) [2, 3, 11, 17, 24, 26, 28]. For such methods to be effective when solving ordinary differential equations (ODEs), it is essential to analyze both their stability and convergence properties. Convergence analysis primarily concerns the behavior of numerical solutions as the step size h approaches zero, while stability analysis examines the range of step sizes h that yield quantitatively or qualitatively correct results without undesirable error growth. A numerical method is convergent if it is both consistent and zero-stable [8, 15]. Consistency relates to the local truncation error of the method, whereas zero-stability ensures that errors remain bounded throughout the computation as the solution progresses [10, 20].

For nonstiff ODEs, establishing convergence may be sufficient. However, for stiff ODEs, convergence alone does not guarantee correct solutions; it is also necessary to determine the region of absolute stability to fully assess the method's performance. In the literature, two main approaches are used to investigate stability: The stability polynomial $\pi(r, z)$ and the stability function $L(z)$. The stability polynomial $\pi(r, z)$ is an algebraic equation in r and z, where $z = h\lambda$ (with h the step size and λ the eigenvalue of the test problem) and r represents the roots of the characteristic equation. For absolute stability, these roots must satisfy $|r_n(z)| < 1$ for $n = 1, 2, \ldots, k$ for a fixed z. The stability function $L(z)$, on the other hand, is a complex-valued function of $z = x + iy$ (with $x, y \in \mathbb{R}$) that represents the spectral radius of the amplification matrix derived from the block method. This amplification matrix can be constructed from the method's defining matrices.

The stability polynomial $\pi(r, z)$ is a tool for analyzing how errors propagate over time and how the choice of step size h influences the method's behavior through the parameter $z = h\lambda$. Proper analysis of the roots r of the stability polynomial helps determine the region of absolute stability, ensuring that the step size h remains within this region to maintain stability and prevent

solution oscillations or blow-up. To derive the stability polynomial for a numerical method, one must obtain its associated characteristic polynomials. In contrast, the stability function $L(z)$ for $z = x+iy$ is a complex-valued function representing the dominant eigenvalue of the amplification matrix [6, 18]. It is typically derived by applying the method to a test problem, and it quantifies how errors are amplified after one step. This function helps determine whether errors grow or decay over time: If $|L(z)| < 1$, then errors decay; if $|L(z)| = 1$, then the method is neutrally stable (errors neither grow nor decay); and if $|L(z)| > 1$, then the method is unstable [13].

A key distinction between the stability polynomial and the stability function lies in their error analysis procedures. The stability polynomial focuses on the locations of the roots r of the characteristic equation in the complex r-plane; for absolute stability, all roots must lie inside the unit disk ($|r| < 1$) when z belongs to the left half-plane. The stability function, however, is concerned with the behavior of $L(z)$ in the complex z-plane.

Akinfenwa, Jator, and Yao [4] investigated the stability of continuous block methods with step numbers $k = 2$ and $k = 3$ for solving stiff problems. They plotted stability regions and the zeros and poles of the stability functions $L(z)$ derived from these methods, concluding that the methods were A-stable based primarily on the absence of poles in the left half-plane. Although they obtained zeros and poles from the stability functions, they did not mention using the stability polynomial—though defined in the work—to plot the regions of absolute stability. Their study was limited to block backward differentiation formulas (BBDFs). Akinnukawe, Akinfenwa, and Okunuga [5] employed a similar approach, deriving the stability function $L(z)$ for a block numerical integrator and showing it to be A-stable and L-stable, but they presented the stability region plot without indicating the positions of the zeros and poles. Their work also focused exclusively on BBDF methods, without considering other classes of block numerical integrators. Other researchers, such as Ibrahim, Othman, and Suleiman [16], have also focused on BBDF methods, presenting stability regions without explicitly showing zeros and poles in the complex plane.

Ngwane and Jator [18] plotted zeros and poles from the stability function of a derived block method while determining its stability region; however, they did not attribute their plot to the stability polynomial. In the context of multistep block methods for ODEs, Chu and Hamilton [7] analyzed the stability of finite difference equations using vectorial representation. Their analysis did not explicitly define a stability function, but their elegant stability analysis of predictor-corrector methods provided insights into numerical schemes that, when applied to a test equation, produce a coefficient matrix whose spectral radius is less than unity—a condition for absolute stability of block methods. A similar procedure can be found in Celaya and Anza [6].

Prior studies on block methods [1, 16, 17, 18, 25] have typically used stability polynomials or stability functions in isolation, without reconciling the two approaches or addressing whether their stability conclusions must coincide. What some researchers call the stability function in one variable z [2] is what others call the stability polynomial [23]. However, many discrepancies exist in the literature: The stability polynomial $\pi(r, z)$ is an algebraic equation in r and z; the stability function $L(z)$, on the other hand, is a complex-valued function of z only. While $z = h\lambda$ in $\pi(r, z)$, it is $z = x + iy$ in $L(z)$. The representation of z in both $\pi(r, z)$ and $L(z)$ is not exactly the same. While z is used in $\pi(r, z)$, the objective is to obtain eigenvalues $r_i(z)$ of the amplification matrix $D(z)$ (roots of $\det(rI - D(z)) = 0$); however, the use of z in the stability function $L(z)$ aims at obtaining zeros and poles of $L(z)$ and their positions in the complex plane. The motivation for this study is to investigate whether the stability polynomial and the stability function yield identical A-stability conclusions for consistent methods, thereby providing a novel theoretical verification mechanism. This paper's novelty is threefold: (i) proving that for consistent block methods both tools must yield congruent stability conclusions, with any discrepancy indicating derivation errors; (ii) demonstrating that the absence of poles in the left half-plane alone does not guarantee A-stability, thereby correcting a common misconception in the literature; and (iii) offering practical guidance on tool selection—$\pi(r, z)$ is preferable for determining $A(\alpha)$-stability angles, while $L(z)$ is better suited for finding zeros and poles, and for L-stability analysis (a much-desired stability property) due to its rational form. Unlike prior derivation-focused works, this study provides a unified comparative framework. For inconsistent methods (which are not convergent), the question of A-stability is not meaningful. Discrepancies between the two approaches for consistent methods serve as theoretical indicators of potential derivation errors or method inconsistencies. This study also addresses the research question: Does the absence of poles in the left half-plane of the complex plane guarantee that a block method is A-stable, as suggested in some literature? It is important to note that deriving block hybrid methods and applying them to solve stiff ODEs without a detailed analysis of the stability region risks selecting an inappropriate method for very stiff problems, leading to incorrect solutions. Determining stability regions with respect to a test problem allows for the selection of an appropriate step size that maintains both accuracy and efficiency while ensuring numerical stability for specific ODEs. Furthermore, some researchers have derived $A(\alpha)$-stable or almost A-stable block methods without specifying the angle α [16], which limits the applicability of such methods to stiff problems. For instance, Ibrahim, Othman, and Suleiman [16] developed an implicit $r = 3$ BBDF which they referred to as an almost A-stable method and, via a plot, showed $A(\alpha)$-stability without specifying the angle α. Knowing the value of α is crucial when considering these methods for solving stiff ODEs. A key theoretical contribution of this work is the demonstration that for consistent block methods, stability polynomials and stability functions

must yield congruent stability conclusions; any discrepancy between them serves as a theoretical indicator of potential derivation errors or method inconsistency, providing a novel verification mechanism.

In this paper, we investigate in detail the use of the stability polynomial and stability function for a block hybrid explicit method (BHEM) and two block hybrid implicit methods (BHIM). Further experiments are conducted to explore the strengths and weaknesses of both approaches for other types of block methods, showing the applicability of this research to both block hybrid and nonhybrid block methods. For illustration, we compare the accuracy of the derived block methods with MATLAB's `ode23s` solver. The remainder of this paper is organized as follows: Section 2 presents the derivation of the block hybrid methods. Section 3 analyzes the stability polynomials and stability functions of the new methods and includes their plots. Section 4 describes the implementation of the proposed methods, and Section 5 presents and discusses the results. Finally, section 6 provides the conclusions.

2 Derivation of the block methods

The methodology for deriving the block hybrid methods in this study follows the approach of Onumanyi, Sirisena, and Jator [22]. We derive three block hybrid methods from continuous k-step linear multistep methods. To begin, we define a polynomial of degree $p = t+m-1$, where $t > 0$ and $m > 0$, as

$$y(x) = \sum_{j=0}^{t-1} \alpha_j(x) y(x_{n+j}) + h \sum_{j=0}^{m-1} \beta_j(x) f(x_j, y(x_j)), \quad x_n \leqslant x \leqslant x_{n+k}, \tag{1}$$

where $\alpha_j(x)$ and $\beta_j(x)$ are continuous coefficients defined by

$$\alpha_j(x) = \sum_{i=0}^{t+m-1} \alpha_{j,i+1} x^i, \quad j \in \{0, \ldots, t-1\} \tag{2}$$

and

$$h\beta_j(x) = h \sum_{i=0}^{t+m-1} \beta_{j,i+1} x^i, \quad j \in \{0, \ldots, m\}. \tag{3}$$

In (1), the points x_j for $j = 0, \ldots, m-1$ are the m distinct collocation points, while the points x_{n+j} for $j = 0, \ldots, t-1$ (with $0 < t \leqslant k$) are the interpolation points.

The continuous coefficients $\alpha_j(x)$ and $\beta_j(x)$ can be obtained by inverting the multistep collocation matrix D, defined as

$$D = \begin{bmatrix} 1 & x_n & x_n^2 & \cdots & x_n^{t+m-1} \\ 1 & x_{n+1} & x_{n+1}^2 & \cdots & x_{n+1}^{t+m-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n+t-1} & x_{n+t-1}^2 & \cdots & x_{n+t-1}^{t+m-1} \\ 0 & 1 & 2x_0 & \cdots & (t+m-1)x_0^{t+m-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & 2x_{m-1} & \cdots & (t+m-1)x_{m-1}^{t+m-1} \end{bmatrix}. \tag{4}$$

The dimension of D is $(t+m)\times(t+m)$. Let $C = D^{-1}$. Through simple matrix operations, we obtain

$$DC = I, \tag{5}$$

where I is the identity matrix of dimension $(t+m)$. For a more detailed explanation of multistep collocation using the matrix inversion technique, readers are referred to Onumanyi, Sirisena, and Jator [22]. We now proceed to derive a BHEM and two BHIMs.

2.1 Continuous formulation of a two-step explicit method incorporating one off-grid collocation point

Let $k = 2$, let $t = 2$, and let $m = 2$, and let us incorporate the off-grid collocation point $x_{n+\frac{49}{25}}$. Then (1) becomes

$$y(x) = \alpha_0(x)y_n + \alpha_1(x)y_{n+1} + h\left[\beta_1(x)f_{n+1} + \beta_{\frac{49}{25}}(x)f_{n+\frac{49}{25}}\right], \tag{6}$$

and the matrix D from (4) becomes

$$D = \begin{bmatrix} 1 & x_n & x_n^2 & x_n^3 \\ 1 & x_{n+1} & x_{n+1}^2 & x_{n+1}^3 \\ 0 & 1 & 2x_{n+1} & 3x_{n+1}^2 \\ 0 & 1 & 2x_{n+\frac{49}{25}} & 3x_{n+\frac{49}{25}}^2 \end{bmatrix}. \tag{7}$$

Using computational software to compute $C = D^{-1}$ yields the continuous formulation for (6) as

$$\begin{aligned} y(x) = & \frac{(h-x+x_n)^2}{h^3}\left(h - \frac{25x}{61} + \frac{25x_n}{61}\right) y_n \\ & + \frac{147(x-x_n)}{61h^3}\left(h^2 + \left(-\frac{37x}{49} + \frac{37x_n}{49}\right)h + \frac{25(x-x_n)^2}{147}\right) y_{n+1} \\ & - \frac{(-1825x + 1825x_n + 4753h)(x-x_n)(h-x+x_n)}{2928h^2} f_{n+1} \\ & + \frac{(625x - 625x_n)(h-x+x_n)^2}{2928h^2} f_{n+\frac{49}{25}}. \end{aligned} \tag{8}$$

Evaluating (8) at $x = x_{n+1}$, $x = x_{n+2}$, and its derivative at $x = x_n$ yields

$$\left.\begin{aligned} y_{n+1} &= y_n + \frac{341}{144}hf_{n+1} - \frac{3125}{144}hf_{n+\frac{49}{25}} + \frac{61}{3}hf_{n+2}, \\ y_{n+\frac{49}{25}} &= \frac{6912}{38125}y_n + \frac{31213}{38125}y_{n+1} + \frac{28812}{38125}hf_{n+1} + \frac{588}{1525}hf_{n+\frac{49}{25}}, \\ y_{n+2} &= \frac{11}{61}y_n + \frac{50}{61}y_{n+1} + \frac{1103}{1464}hf_{n+1} + \frac{625}{1464}hf_{n+\frac{49}{25}}. \end{aligned}\right\} \tag{9}$$

2.2 Continuous formulation of a two-step implicit method incorporating one off-grid interpolation point

Let $k = 2$, $t = 2$, $m = 3$, and let $x = x_{n+\frac{125}{64}}$ be an off-grid interpolation point. Then (1) becomes

$$y(x) = \alpha_1(x)y_{n+1} + \alpha_{\frac{125}{64}}(x)y_{n+\frac{125}{64}} + h\left[\beta_0(x)f_n + \beta_1(x)f_{n+1} + \beta_2(x)f_{n+2}\right], \tag{10}$$

and the matrix D from (4) becomes

$$D = \begin{pmatrix} 1 & x_{n+1} & x_{n+1}^2 & x_{n+1}^3 & x_{n+1}^4 \\ 1 & x_{n+\frac{125}{64}} & x_{n+\frac{125}{64}}^2 & x_{n+\frac{125}{64}}^3 & x_{n+\frac{125}{64}}^4 \\ 0 & 1 & 2x_n & 3x_n^2 & 4x_n^3 \\ 0 & 1 & 2x_{n+1} & 3x_{n+1}^2 & 4x_{n+1}^3 \\ 0 & 1 & 2x_{n+2} & 3x_{n+2}^2 & 4x_{n+2}^3 \end{pmatrix}. \tag{11}$$

Using computational software to obtain $C = D^{-1}$ and substituting into (10) yields the continuous formulation

$$
\begin{aligned}
y(x) = &\left[\left(-\frac{140625}{16636591h^4}\left(\frac{64x_n}{3}+h-\frac{64x}{3}\right)\right)\right.\\
&\times\left(h^2+\frac{8192}{375}(x-x_n)h-\frac{4096}{375}(x-x_n)^2\right)\\
&\left.\times\left(\frac{64x_n}{125}+h-\frac{64x}{125}\right)\right]y_{n+1}\\
&+\left[\frac{\left(16777216h^2+16777216(2x-2x_n)h-16777216(x-x_n)^2\right)(h-x+x_n)^2}{16636591h^4}\right]y_{n+\frac{125}{64}}\\
&+\left[-\frac{(h-x+x_n)^2(64x_n+125h-64x)(-70x+70x_n+143h)}{53652h^3}\right]f_n\\
&+\left[\left(-\frac{1093750(h-x+x_n)}{818193h^3}\left(\frac{64x_n}{125}+h-\frac{64x}{125}\right)\right)\right.\\
&\left.\times\left(h^2+\frac{189}{125}(x-x_n)h-\frac{8567}{8750}(x-x_n)^2\right)\right]f_{n+1}\\
&+\left[-\frac{(64x_n+125h-64x)(314x-314x_n+125h)(h-x+x_n)^2}{53652h^3}\right]f_{n+2}. \qquad (12)
\end{aligned}
$$

Evaluating (12) at $x = x_n$, $x = x_{n+2}$, and differentiating the continuous scheme (12) and evaluating it at $x = x_{n+\frac{125}{64}}$ yields

$$
\begin{aligned}
y_{n+1} =& -\frac{16636591}{140625}y_n+\frac{16777216}{140625}y_{n+\frac{125}{64}}-\frac{532103}{13500}hf_n-\frac{4270}{27}hf_{n+1}\\
&-\frac{3721}{108}hf_{n+2},\\
y_{n+\frac{125}{64}} =& y_{n+1}-\frac{15207727}{786432000}hf_n+\frac{1272155}{3145728}hf_{n+1}+\frac{272731}{96000}hf_{n+\frac{125}{64}}\\
&-\frac{4766601}{2097152}hf_{n+2},\\
y_{n+2} =& -\frac{140625}{16636591}y_{n+1}+\frac{16777216}{16636591}y_{n+\frac{125}{64}}+\frac{3}{17884}hf_n-\frac{942}{272731}hf_{n+1}\\
&+\frac{753}{17884}hf_{n+2}. \qquad (13)
\end{aligned}
$$

2.3 Continuous formulation of a three-step implicit method incorporating one off-grid collocation point

Let $k = 3$, let $t = 3$, let $m = 2$, and let $x_{n+\frac{72}{25}}$ be an off-grid collocation point. Then (1) becomes

$$
y(x) = \alpha_0(x)y_n+\alpha_1(x)y_{n+1}+\alpha_2(x)y_{n+2}+h\left[\beta_{\frac{72}{25}}(x)f_{n+\frac{72}{25}}+\beta_3(x)f_{n+3}\right], \qquad (14)
$$

and the matrix D from (4) becomes

$$D = \begin{pmatrix} 1 & x_n & x_n^2 & x_n^3 & x_n^4 \\ 1 & x_{n+1} & x_{n+1}^2 & x_{n+1}^3 & x_{n+1}^4 \\ 1 & x_{n+2} & x_{n+2}^2 & x_{n+2}^3 & x_{n+2}^4 \\ 0 & 1 & 2x_{n+\frac{72}{25}} & 3x_{n+\frac{72}{25}}^2 & 4x_{n+\frac{72}{25}}^3 \\ 0 & 1 & 2x_{n+3} & 3x_{n+3}^2 & 4x_{n+3}^3 \end{pmatrix}. \tag{15}$$

Using computational software to obtain $C = D^{-1}$ and substituting into (14) yields

$$\begin{aligned} y(x) =& \frac{1}{h^4}(x_n + h - x)\left(\frac{1}{2}x_n + h - \frac{1}{2}x\right) \\ &\times \left(h^2 + \left(-\frac{54033x}{93146} + \frac{54033x_n}{93146}\right)h + \frac{8075(x - x_n)^2}{93146}\right) y_n \\ &+ \frac{1}{46573h^4}(181440x - 181440x_n)\left(\frac{1}{2}x_n + h - \frac{1}{2}x\right) \\ &\times \left(h^2 + \left(-\frac{2879x}{5040} + \frac{2879x_n}{5040}\right)h + \frac{1535(x - x_n)^2}{18144}\right) y_{n+1} \\ &- \frac{(84564x - 84564x_n)(x_n + h - x)}{46573h^4} \\ &\times \left(h^2 + \left(-\frac{6215x}{12528} + \frac{6215x_n}{12528}\right)h + \frac{22625(x - x_n)^2}{338256}\right) y_{n+2} \\ &+ \frac{(15625x - 15625x_n)(x_n + 2h - x)(11x_n + 39h - 11x)(x_n + h - x)}{279438h^3} f_{n+\frac{72}{25}} \\ &- \frac{(x_n + 2h - x)(x - x_n)(75025x_n + 253296h - 75025x)(x_n + h - x)}{139719h^3} f_{n+3}. \end{aligned} \tag{16}$$

Evaluating (16) at $x = x_{n+\frac{72}{25}}$, $x = x_{n+3}$, and differentiating the continuous scheme (14) with evaluation at $x = x_{n+1}$ and $x = x_{n+2}$ yield

$$\begin{aligned} y_{n+1} =& -\frac{251}{776}y_n + \frac{1027}{776}y_{n+2} - \frac{46573}{36472}hf_{n+1} - \frac{109375}{54708}hf_{n+\frac{72}{25}} \\ &+ \frac{3793}{2328}hf_{n+3}, \\ y_{n+2} =& -\frac{395}{2837}y_n + \frac{3232}{2837}y_{n+1} + \frac{46573}{31207}hf_{n+2} - \frac{265625}{93621}hf_{n+\frac{72}{25}} \\ &+ \frac{18772}{8511}hf_{n+3}, \\ y_{n+\frac{72}{25}} =& \frac{29134501}{727703125}y_n - \frac{203233536}{727703125}y_{n+1} + \frac{180360432}{145540625}y_{n+2} \\ &+ \frac{2270664}{1164325}hf_{n+\frac{72}{25}} - \frac{923750784}{727703125}hf_{n+3} \\ y_{n+3} =& \frac{1861}{46573}y_n - \frac{12987}{46573}y_{n+1} + \frac{57699}{46573}y_{n+2} + \frac{93750}{46573}hf_{n+\frac{72}{25}} \\ &- \frac{56442}{46573}hf_{n+3}. \end{aligned} \tag{17}$$

Equations (9), (13), and (17) represent the derived block hybrid schemes. Their convergence analyses are presented in the next section.

3 Analyses of the derived block methods

This section presents convergence analyses and plots of the regions of absolute stability for the derived block methods.

3.1 Convergence analysis of the block method (9)

We employ the approach of Fatunla [12] by reformulating (9) as

$$M^{(1)}Y_{n+i} = M^{(0)}Y_n + hB^{(1)}f_{n+i}, \quad i \in \mathbb{R}, \tag{18}$$

where

$$M^{(1)} = \begin{pmatrix} 1 & 0\ 0 \\ -\frac{31213}{38125} & 1\ 0 \\ -\frac{50}{61} & 0\ 1 \end{pmatrix}, \quad M^{(0)} = \begin{pmatrix} 0\ 0 & 1 \\ 0\ 0 & \frac{6912}{38125} \\ 0\ 0 & \frac{11}{61} \end{pmatrix}, \tag{19}$$

and

$$B^{(1)} = \begin{pmatrix} \frac{341}{144} & -\frac{3125}{144} & \frac{61}{3} \\ \frac{28812}{38125} & \frac{588}{1525} & 0 \\ \frac{1103}{1464} & \frac{625}{1464} & 0 \end{pmatrix}. \tag{20}$$

The first characteristic polynomial for (9) is given by

$$\rho(\lambda) = \det(\lambda M^{(1)} - M^{(0)}) = |\lambda M^{(1)} - M^{(0)}| = 0.$$

Thus, we obtain

$$\rho(\lambda) = \left| \lambda \begin{pmatrix} 1 & 0\ 0 \\ -\frac{31213}{38125} & 1\ 0 \\ -\frac{50}{61} & 0\ 1 \end{pmatrix} - \begin{pmatrix} 0\ 0 & 1 \\ 0\ 0 & \frac{6912}{38125} \\ 0\ 0 & \frac{11}{61} \end{pmatrix} \right| = \lambda^2(\lambda - 1),$$

which implies

$$\lambda_1 = 1, \quad \lambda_2 = 0, \quad \lambda_3 = 0. \tag{21}$$

Since $|\lambda_i| \leqslant 1$ for $i = 1, 2, 3$, the block method (9) is zero-stable by Fatunla [11], with order $p = 3$ and error constants $\left(-\frac{83}{360}, -\frac{691488}{23828125}, -\frac{527}{18300}\right)$. Since (9) is zero-stable and has order $p > 1$, it is convergent by Henrici [15].

3.2 Convergence analysis of the block method (13)

We reformulate (13) as

$$M^{(1)}Y_{n+i} = M^{(0)}Y_n + h\left(B^{(1)}f_{n+i} + B^{(0)}f_n\right), \quad i \in \mathbb{R}, \tag{22}$$

where

$$M^{(1)} = \begin{pmatrix} 1 & -\frac{16777216}{140625} & 0 \\ -1 & 1 & 0 \\ \frac{140625}{16636591} & -\frac{16777216}{16636591} & 1 \end{pmatrix}, \quad M^{(0)} = \begin{pmatrix} 0 & 0 & -\frac{16636591}{140625} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{23}$$

and

$$B^{(1)} = \begin{pmatrix} -\frac{4270}{27} & 0 & -\frac{3721}{108} \\ \frac{1272155}{3145728} & \frac{272731}{96000} & -\frac{4766601}{2097152} \\ -\frac{942}{272731} & 0 & \frac{753}{17884} \end{pmatrix}, \quad B^{(0)} = \begin{pmatrix} 0 & 0 & -\frac{532103}{13500} \\ 0 & 0 & -\frac{15207727}{786432000} \\ 0 & 0 & \frac{3}{17884} \end{pmatrix}. \tag{24}$$

The first characteristic polynomial for (13) is

$$\rho(\lambda) = \det(\lambda M^{(1)} - M^{(0)}) = |\lambda M^{(1)} - M^{(0)}| = 0. \tag{25}$$

Thus, we obtain

$$\begin{aligned} \rho(\lambda) &= \left| \lambda \begin{pmatrix} 1 & -\frac{16777216}{140625} & 0 \\ -1 & 1 & 0 \\ \frac{140625}{16636591} & -\frac{16777216}{16636591} & 1 \end{pmatrix} - \begin{pmatrix} 0 & 0 & -\frac{16636591}{140625} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right| \\ &= -\frac{16636591}{140625}\lambda^3 + \frac{16636591}{140625}\lambda^2, \end{aligned}$$

which implies

$$\lambda_1 = 1, \quad \lambda_2 = 0, \quad \lambda_3 = 0. \tag{26}$$

Since $|\lambda_i| \leqslant 1$ for $i = 1, 2, 3$, the block method (13) is zero-stable by Fatunla [11], with order $p = 3$ and error constants

$$\left(\frac{271633}{207360}, \frac{6763352857}{1546188226560}, -\frac{439}{11445760}\right).$$

By Henrici [15], the block method (13) has order $p > 1$ and is zero-stable; therefore, it is convergent.

3.3 Convergence analysis of the block method (17)

We reformulate (17) as

$$M^{(1)} Y_{n+i} = M^{(0)} Y_n + hB^{(1)} f_{n+i}, \quad i \in \mathbb{R}, \tag{27}$$

where

$$M^{(1)} = \begin{pmatrix} 1 & -\frac{1027}{776} & 0 & 0 \\ -\frac{3232}{2837} & 1 & 0 & 0 \\ \frac{203233536}{727703125} & -\frac{180360432}{145540625} & 1 & 0 \\ \frac{12987}{46573} & -\frac{57699}{46573} & 0 & 1 \end{pmatrix}, \quad M^{(0)} = \begin{pmatrix} 0 & 0 & 0 & -\frac{251}{776} \\ 0 & 0 & 0 & -\frac{395}{2837} \\ 0 & 0 & 0 & \frac{29134501}{727703125} \\ 0 & 0 & 0 & \frac{1861}{46573} \end{pmatrix} \tag{28}$$

and

$$B^{(1)} = \begin{pmatrix} -\frac{46573}{36472} & 0 & -\frac{109375}{54708} & \frac{3793}{2328} \\ 0 & \frac{46573}{31207} & -\frac{265625}{93621} & \frac{18772}{8511} \\ 0 & 0 & \frac{2270664}{1164325} & -\frac{923750784}{727703125} \\ 0 & 0 & \frac{93750}{46573} & -\frac{56442}{46573} \end{pmatrix}. \tag{29}$$

The first characteristic polynomial for (17) is

$$\rho(\lambda) = \det(\lambda M^{(1)} - M^{(0)}) = |\lambda M^{(1)} - M^{(0)}| = 0. \tag{30}$$

Thus, we obtain

$$\begin{aligned}\rho(\lambda) &= \left| \lambda \begin{pmatrix} 1 & -\frac{1027}{776} & 0 & 0 \\ -\frac{3232}{2837} & 1 & 0 & 0 \\ \frac{203233536}{727703125} & -\frac{180360432}{145540625} & 1 & 0 \\ \frac{12987}{46573} & -\frac{57699}{46573} & 0 & 1 \end{pmatrix} - \begin{pmatrix} 0 & 0 & 0 & -\frac{251}{776} \\ 0 & 0 & 0 & -\frac{395}{2837} \\ 0 & 0 & 0 & \frac{29134501}{727703125} \\ 0 & 0 & 0 & \frac{1861}{46573} \end{pmatrix} \right| \\ &= -\frac{139719}{275189}\lambda^4 + \frac{139719}{275189}\lambda^3,\end{aligned}$$

which implies

$$\lambda_1 = 1, \quad \lambda_2 = 0, \quad \lambda_3 = 0, \quad \lambda_4 = 0. \tag{31}$$

Since $|\lambda_i| \leqslant 1$ for $i = 1, 2, 3, 4$, the block method (17) is zero-stable by Fatunla [11], with order $p = 4$ and error constants

$$\left(-\frac{863}{207360}, -\frac{20699}{425550}, \frac{38451126384}{2274072265625}, \frac{39267}{2328650}\right).$$

By Henrici [15], (17) has order $p > 1$ and is zero-stable; therefore, it is convergent.

3.4 Stability polynomial of the block method (9)

Definition 1. A block method among (9), (13), or (17) is absolutely stable with region of absolute stability S if all roots r_i of the stability polynomial

$$\pi(r, z) = \rho(r) - z\sigma(r) = 0 \tag{32}$$

satisfy $|r_i| < 1$ for $i = 1, 2, \ldots, k$ for a fixed $z = h\lambda$, where $\rho(r)$ and $\sigma(r)$ are the first and second characteristic polynomials, respectively [4, 21].

Definition 2. A numerical integration scheme is $A(\alpha)$-stable for some $\alpha \in \left[0, \frac{\pi}{2}\right]$ if the wedge

$$S_\alpha = \{z : |\mathrm{Arg}(-z)| < \alpha, z \neq 0\} \tag{33}$$

lies within its region of absolute stability. The largest $\alpha_{\max}$ is called the angle of absolute stability or argument of stability [19].

For block method (9), the stability polynomial $\pi(r,z)$ from (32) is

$$\begin{aligned}\pi\,(r,z) &= M^{(1)}Y_{n+i} - M^{(0)}Y_n - hB^{(1)}f_{n+i} \quad i \in \mathbb{R},\\ &= M^{(1)}r^1 - M^{(0)}r^0 - zB^{(1)}r^1\\ &= M^{(1)}r - M^{(0)} - zB^{(1)}r\\ &= r(M^{(1)} - zB^{(1)}) - M^{(0)}. \end{aligned} \tag{34}$$

Using the values of $M^{(1)}$, $B^{(1)}$, and $M^{(0)}$ from (19) and (20) with computational software, we obtain

$$\pi(r,z) = \begin{pmatrix} r\left(1-\frac{341z}{144}\right) & \frac{3125rz}{144} & -\frac{61rz}{3}-1\\ r\left(-\frac{31213}{38125}-\frac{28812z}{38125}\right) & r\left(1-\frac{588z}{1525}\right) & -\frac{6912}{38125}\\ r\left(-\frac{50}{61}-\frac{1103z}{1464}\right) & -\frac{625rz}{1464} & r-\frac{11}{61}\end{pmatrix}. \tag{35}$$

The determinant of (35) gives the stability polynomial:

$$\pi(r,z) = r^3 - r^2 - \frac{124}{75}r^3z - \frac{26}{75}r^2z + \frac{197}{150}r^3z^2 - \frac{1}{150}r^2z^2 - \frac{49}{75}r^3z^3. \tag{36}$$

The plot of (36) using software is shown in Figure 1.

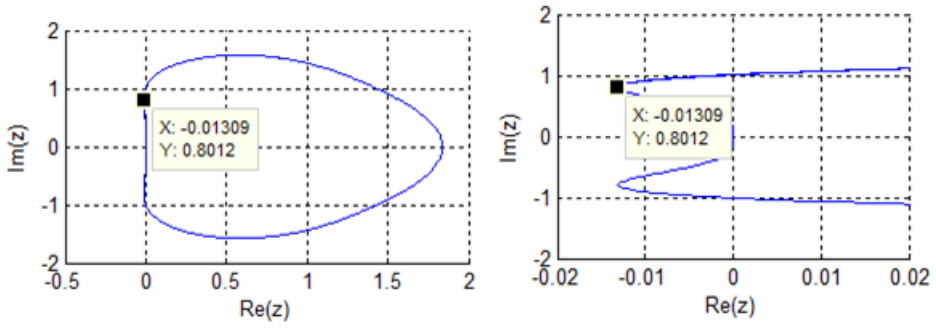


Figure 1: a) (Left) Plot of stability polynomial (36); b) (Right) Detail of (36) near the origin.

3.5 Stability polynomial of the block method (13)

For block method (13), the stability polynomial $\pi(r,z)$ from (32) is

$$\begin{aligned}\pi(r,z) &= M^{(1)}Y_{n+i} - M^{(0)}Y_n - hB^{(1)}f_{n+i} - hB^{(0)}f_n, \quad i \in \mathbb{R},\\ &= M^{(1)}r^1 - M^{(0)}r^0 - zB^{(1)}r^1 - zB^{(0)}r^0\\ &= M^{(1)}r - M^{(0)} - zB^{(1)}r - zB^{(0)}\\ &= r(M^{(1)} - zB^{(1)}) - M^{(0)} - zB^{(0)}. \end{aligned} \tag{37}$$

Using the values of $M^{(1)}$, $B^{(1)}$, $M^{(0)}$, and $B^{(0)}$ from (23) and (24) with computational software, we obtain

$$\pi(r,z) = \begin{pmatrix} r\left(1+\frac{4270\,z}{27}\right) & -\frac{16777216\,r}{140625} & \frac{3721\,rz}{108}+\frac{16636591}{140625}+\frac{532103\,z}{13500} \\ r\left(-1-\frac{1272155\,z}{3145728}\right) & r\left(1-\frac{272731\,z}{96000}\right) & \frac{4766601\,rz}{2097152}+\frac{15207727\,z}{786432000} \\ r\left(\frac{140625}{16636591}+\frac{942\,z}{272731}\right) & -\frac{16777216\,r}{16636591} & r\left(1-\frac{753\,z}{17884}\right)-\frac{3\,z}{17884} \end{pmatrix}. \tag{38}$$

The determinant of (38) gives the stability polynomial:

$$\begin{aligned}\pi(r,z) = &-\frac{16636591}{140625}r^3 + \frac{5273799347}{36000000}r^3z + \frac{216275683}{2400000}r^2z - \frac{8368205273}{108000000}r^3z^2\\ &+ \frac{2279212967}{108000000}r^2z^2 + \frac{16636591}{864000}r^3z^3 + \frac{16636591}{36000000}r^2z^3 + \frac{16636591}{140625}r^2. \end{aligned} \tag{39}$$

The plot of (39) is shown in Figure 2.

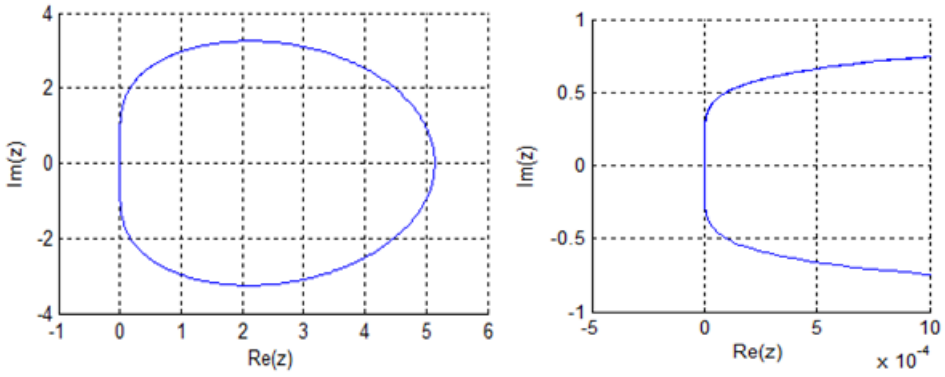


Figure 2: a) (Left) Plot of stability polynomial (39); b) (Right) Detail of (39) near the origin.

3.6 Stability polynomial of the block method (17)

For block method (17), the stability polynomial $\pi(r,z)$ from (32) is

$$\pi\left(r,z\right)=r\left(M^{(1)}-zB^{(1)}\right)-M^{(0)}. \tag{40}$$

Using the values of $M^{(1)}$, $M^{(0)}$, and $B^{(1)}$ from (28) and (29) with computational software, we obtain

$$\pi(r,z)=\begin{pmatrix} r\left(1+\frac{46573\,z}{36472}\right) & -\frac{1027\,r}{776} & \frac{109375\,rz}{54708} & -\frac{3793\,rz}{2328}+\frac{251}{776} \\ -\frac{3232\,r}{2837} & r\left(1-\frac{46573\,z}{31207}\right) & \frac{265625\,rz}{93621} & -\frac{18772\,rz}{8511}+\frac{395}{2837} \\ \frac{203233536\,r}{727703125} & -\frac{180360432\,r}{145540625} & r\left(1-\frac{2270664\,z}{1164325}\right) & \frac{923750784\,rz}{727703125}-\frac{29134501}{727703125} \\ \frac{12987\,r}{46573} & -\frac{57699\,r}{46573} & -\frac{93750\,rz}{46573} & r\left(1+\frac{56442\,z}{46573}\right)-\frac{1861}{46573} \end{pmatrix}. \tag{41}$$

The determinant of (41) gives the stability polynomial:

$$\begin{aligned}\pi(r,z)=&-\frac{139719}{275189}r^4+\frac{15508809}{13759450}r^4z+\frac{139719}{275189}r^3-\frac{32927111}{27518900}r^4z^2\\&+\frac{5449041}{13759450}r^3z+\frac{21935883}{27518900}r^4z^3+\frac{2747807}{27518900}r^3z^2\\&-\frac{2514942}{6879725}r^4z^4+\frac{139719}{27518900}r^3z^3.\end{aligned} \tag{42}$$

The plot of (42) is shown in Figure 3.

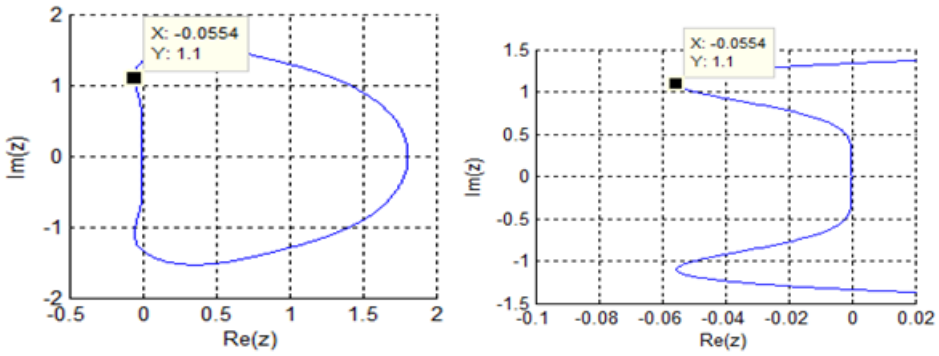


Figure 3: a) (Left) Plot of stability polynomial (42); b) (Right) Detail of (42) near the origin.

3.7 *L*-stability analysis via stability polynomials

Beyond A-stability, L-stability requires that all characteristic roots $r_i(z)$ of the stability polynomial satisfy $|r_i(z)| \to 0$ as $z \to -\infty$ along the real axis. For $L(\alpha)$-stability, this damping must occur within the wedge of $A(\alpha)$-stability. We analyze this asymptotic behavior for each method by examining the highest powers of z in their respective stability polynomials.

Block method (9)**:** The stability polynomial from (36) is

$$\pi(r,z) = r^3 - r^2 - \frac{124}{75}r^3z - \frac{26}{75}r^2z + \frac{197}{150}r^3z^2 - \frac{1}{150}r^2z^2 - \frac{49}{75}r^3z^3.$$

For large negative z, the dominant term is the highest power of z, which is z^3. Retaining only this term, we have

$$\pi(r,z) \approx -\frac{49}{75}r^3z^3 = 0.$$

This implies that all three roots satisfy $r \to 0$ as $z \to -\infty$. Since method (9) is $A(\alpha)$-stable with $\alpha = 89.06°$, it follows that it is $L(\alpha)$-stable within this wedge.

Block method (13)**:** The stability polynomial from (39) is

$$\begin{aligned}\pi(r,z) = &-\frac{16636591}{140625}r^3 + \frac{5273799347}{36000000}r^3z + \frac{216275683}{2400000}r^2z \\ &- \frac{8368205273}{108000000}r^3z^2 + \frac{2279212967}{108000000}r^2z^2 + \frac{16636591}{864000}r^3z^3 \\ &+ \frac{16636591}{36000000}r^2z^3 + \frac{16636591}{140625}r^2.\end{aligned}$$

The highest power of z is z^3. Isolating the z^3 terms, we have

$$\left(\frac{16636591}{864000}r^3 + \frac{16636591}{36000000}r^2\right)z^3 = 0.$$

Setting the coefficient to zero yields

$$r^2\left(\frac{1}{864000}r + \frac{1}{36000000}\right) = 0.$$

Thus, two roots approach $r = 0$, and the third root approaches $r = -\frac{864000}{36000000} = -0.024 \neq 0$ as $z \to -\infty$. Since not all roots tend to zero, method (13) is not L-stable, despite being A-stable.

Block method (17)**:** The stability polynomial from (42) is

$$\begin{aligned}\pi(r,z) = &-\frac{139719}{275189}r^4 + \frac{15508809}{13759450}r^4z + \frac{139719}{275189}r^3 - \frac{32927111}{27518900}r^4z^2 \\ &+ \frac{5449041}{13759450}r^3z + \frac{21935883}{27518900}r^4z^3 + \frac{2747807}{27518900}r^3z^2\end{aligned}$$

$$-\frac{2514942}{6879725}r^4z^4+\frac{139719}{27518900}r^3z^3.$$

The highest power of z is z^4. The dominant term is

$$-\frac{2514942}{6879725}r^4z^4=0,$$

which implies all four roots satisfy $r\to 0$ as $z\to-\infty$. Given that method (17) is $A(\alpha)$-stable with $\alpha=87.12^\circ$, it is $L(\alpha)$-stable within this wedge.

This analysis demonstrates that while stability polynomials are primarily used to determine A-stability regions, their asymptotic structure can also reveal L-stability properties by examining the behavior of roots as $z\to-\infty$.

3.8 Stability function of the block method (9)

Definition 3 (A-Stability using stability function)**.** Let $L(z)$ be the stability function defined by $L(z)=\frac{P(z)}{Q(z)}$, where $P(z)$ has degree k and $Q(z)$ has degree j. A method with stability function $L(z)$ is A-stable if and only if

$$|L(iy)|\leqslant 1,\quad \text{for all real } y, \tag{43}$$

and

$$L(z)\ \text{ is analytic for } \Re(z)<0. \tag{44}$$

Condition (43) alone, known as I-stability, ensures stability on the imaginary axis. It is equivalent to the condition that the polynomial

$$E(y)=|Q(iy)|^2-|P(iy)|^2=Q(iy)\,Q(-iy)-P(iy)\,P(-iy) \tag{45}$$

satisfies

$$E(y)\geq 0\quad \text{for all } y. \tag{46}$$

(see [14]).

Remark 1. The above definition from Hairer and Wanner [14] applies to one-step methods, where the stability function $L(z)$ completely characterizes stability. For block methods, $L(z)$ is

typically taken as the spectral radius $\rho(D(z))$ of the amplification matrix. Conditions (43) and (44) are necessary but not sufficient for A-stability. The insufficiency arises because $L(z)$ captures only the dominant eigenvalue; nondominant eigenvalues may violate the stability requirements even when the spectral radius satisfies the conditions. The necessary and sufficient conditions for A-stability of block methods are precisely those stated in Definition 4, which require that *all* eigenvalues of the amplification matrix satisfy the appropriate bounds (with the additional requirement that any eigenvalue on the unit circle be simple). Verifying these conditions is equivalent to checking that all roots of the stability polynomial $\pi(r, z) = 0$ lie in the closed unit disk for $\Re(z) \leq 0$, with strict inequality for $\Re(z) < 0$ and simplicity on the boundary. When nondominant eigenvalues are identically zero or have moduli orders of magnitude smaller than the dominant eigenvalue, conditions (43) and (44) may suffice for practical computations, but they do not constitute a rigorous mathematical proof of A-stability.

Definition 4 (Necessary and sufficient conditions for A-stability of block methods)**.** A consistent block method with amplification matrix $D(z)$ and eigenvalues $\{r_i(z)\}_{i=1}^{k}$ is A-stable if and only if

1. $|r_i(iy)| \leq 1$ for all real y and all $i = 1, \ldots, k$.
2. $|r_i(z)| < 1$ for all z with $\Re(z) < 0$ and all $i = 1, \ldots, k$.
3. any eigenvalue with $|r_i(iy)| = 1$ is simple (to avoid polynomial growth).

Theorem 1 (A-stability verification)**.** A block method is A-stable if and only if

1. All roots r of the stability polynomial $\pi(r, iy) = 0$ satisfy $|r| \leq 1$ for all real y.
2. All roots r of the stability polynomial $\pi(r, z) = 0$ satisfy $|r| < 1$ for all z with $\Re(z) < 0$.

Proof. Let the block method be applied to the test equation $y' = \lambda y$, with $z = h\lambda$. This yields an amplification matrix $D(z)$ such that the numerical solution advances as $Y_{n+1} = D(z)Y_n$. The stability polynomial is defined as $\pi(r, z) = \det(rI - D(z))$. Hence, the roots r of $\pi(r, z) = 0$ are exactly the eigenvalues of $D(z)$.

If the method is A-stable according to Definition 4, then by that definition, we have the following results:

- For every real y, each eigenvalue r of $D(iy)$ satisfies $|r| \leq 1$. This is precisely condition (1) of the theorem, because those eigenvalues are the roots of $\pi(r, iy) = 0$.

- For every z with $\Re(z) < 0$, each eigenvalue r of $D(z)$ satisfies $|r| < 1$. This is exactly condition (2) of the theorem, since those eigenvalues are the roots of $\pi(r, z) = 0$.

Conversely, assume conditions (1) and (2) hold. Then for every real y, all roots of $\pi(r, iy) = 0$ satisfy $|r| \leq 1$, which means all eigenvalues of $D(iy)$ satisfy $|r| \leq 1$. Similarly, for every z with $\Re(z) < 0$, all roots of $\pi(r, z) = 0$ satisfy $|r| < 1$, that is, all eigenvalues of $D(z)$ satisfy $|r| < 1$. Thus the method meets both requirements of Definition 4 and is therefore A-stable.

Hence, the two conditions are necessary and sufficient for A-stability. □

Remark 2. For a certain class of block methods where the amplification matrix $D(z)$ has rank 1 for all z, exactly one eigenvalue is nonzero and the remaining $k-1$ eigenvalues are identically zero. In such cases, the zero eigenvalues trivially satisfy any modulus bound satisfied by the nonzero eigenvalue. Consequently, verifying A-stability reduces to checking only the dominant eigenvalue, provided no poles exist in the left half-plane. However, this property is not general; some block methods may yield amplification matrices of full rank, requiring analysis of all eigenvalues.

To obtain the stability function for block method (9), we apply (18) to the test problem $y' = \lambda y$, $\lambda < 0$:

$$M^{(1)}Y_{n+1} = M^{(0)}Y_n + hB^{(1)}f_{n+1},$$

$$\begin{aligned}
M^{(1)}Y_{n+1} &= M^{(0)}Y_n + hB^{(1)}\lambda Y_{n+1},\\
M^{(1)}Y_{n+1} - h\lambda B^{(1)}Y_{n+1} &= M^{(0)}Y_n,\\
(M^{(1)} - h\lambda B^{(1)})Y_{n+1} &= M^{(0)}Y_n,\\
(M^{(1)} - zB^{(1)})Y_{n+1} &= M^{(0)}Y_n, \quad z = h\lambda,
\end{aligned}$$

$$Y_{n+1} = (M^{(1)} - zB^{(1)})^{-1}M^{(0)}Y_n, \tag{47}$$

$$Y_{n+1} = D(z)Y_n, \tag{48}$$

where $D(z) = (M^{(1)} - zB^{(1)})^{-1}M^{(0)}$ is a matrix with eigenvalues $\{d_1, d_2, \ldots, d_k\} = \{0, 0, \ldots, d_k\}$ such that the dominant eigenvalue d_k is the stability function $L(z) : \mathbb{C} \to \mathbb{C}$, a rational function with real coefficients. Using the values of $M^{(1)}$, $B^{(1)}$, and $M^{(0)}$ from (19) and (20) with Maple software, we obtain

$$L(z) = \frac{-z^2 - 52\,z - 150}{98\,z^3 - 197\,z^2 + 248\,z - 150}. \tag{49}$$

The plot of the stability function (49) is shown in Figure 4.

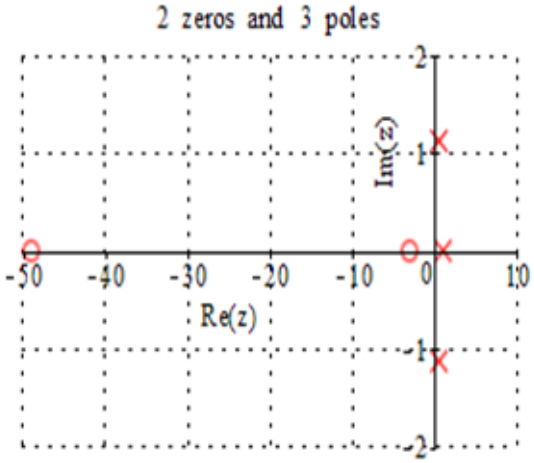


Figure 4: Plot of stability function (49).

3.9 Stability function of the block method (13)

For the stability function of block method (13), we apply (22) to the test problem $y' = \lambda y$, $\lambda < 0$:

$$M^{(1)}Y_{n+1} = M^{(0)}Y_n + hB^{(1)}f_{n+1} + hB^{(0)}f_n,$$

$$\begin{aligned}
M^{(1)}Y_{n+1} &= M^{(0)}Y_n + hB^{(1)}\lambda Y_{n+1} + hB^{(0)}\lambda Y_n,\\
M^{(1)}Y_{n+1} - h\lambda B^{(1)}Y_{n+1} &= M^{(0)}Y_n + h\lambda B^{(0)}Y_n,\\
(M^{(1)} - h\lambda B^{(1)})Y_{n+1} &= (M^{(0)} + h\lambda B^{(0)})Y_n,\\
(M^{(1)} - zB^{(1)})Y_{n+1} &= (M^{(0)} + zB^{(0)})Y_n, \quad z = h\lambda,
\end{aligned}$$

$$Y_{n+1} = (M^{(1)} - zB^{(1)})^{-1}(M^{(0)} + zB^{(0)})Y_n, \tag{50}$$

$$Y_{n+1} = D(z)Y_n, \tag{51}$$

where $D(z) = (M^{(1)} - zB^{(1)})^{-1}(M^{(0)} + zB^{(0)})$. Using the values of $M^{(1)}$, $B^{(1)}$, $M^{(0)}$, and $B^{(0)}$ from (23) and (24) with computational software, we obtain

$$L(z) = \frac{-3\,z^3 - 137\,z^2 - 585\,z - 768}{125\,z^3 - 503\,z^2 + 951\,z - 768}. \tag{52}$$

The plot of the stability function (52) is shown in Figure 5.

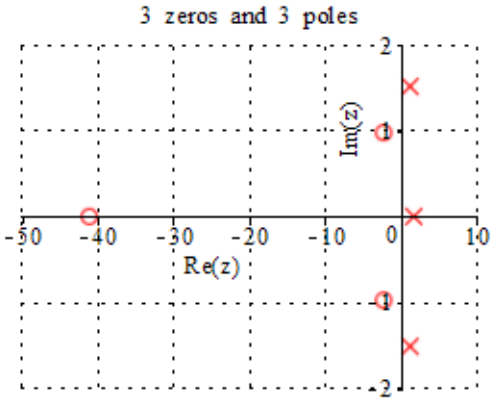


Figure 5: Plot of stability function (52).

3.10 Stability function of the block method (17)

To obtain the stability function for block method (17), we apply (27) to the test problem $y' = \lambda y$, $\lambda < 0$, which gives

$$Y_{n+1} = D(z)Y_n, \tag{53}$$

where $D(z) = (M^{(1)} - zB^{(1)})^{-1}M^{(0)}$. Using the values of $M^{(1)}$, $B^{(1)}$, and $M^{(0)}$ from (28) and (29) with computational software, we obtain

$$L(z) = \frac{3\,z^3 + 59\,z^2 + 234\,z + 300}{216\,z^4 - 471\,z^3 + 707\,z^2 - 666\,z + 300}. \tag{54}$$

The plot of the stability function (54) is shown in Figure 6.

Remark 3. This analysis provides a definitive answer to the research question posed in Section 1: "Does the absence of poles in the left half-plane of the complex plane guarantee that a block

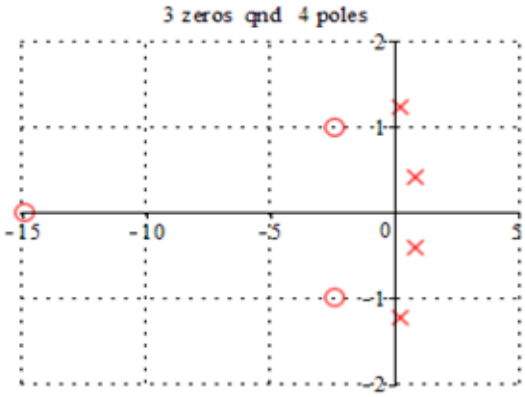


Figure 6: Plot of stability function (54).

method is A-stable, as suggested in some literature?" The answer is no. While the absence of poles in the left half-plane is necessary for A-stability, it is not sufficient. Condition (1) of Definition 4 (boundedness on the imaginary axis) must also be satisfied. Method (9) exemplifies this: it has no poles in the left half-plane but violates condition (1) of Definition 4, rendering it not A-stable. Similarly, method (17) does not satisfy condition (1) of Definition 4. Only method (13) satisfies all required conditions and is A-stable. Therefore, all conditions must be verified when using stability functions to assess A-stability.

3.11 *L*-stability analysis via stability function

Beyond the A-stability and $A(\alpha)$-stability analyzed above, L-stability is a stronger property that ensures the numerical method damps out stiff components rapidly. A method is L-stable if it is A-stable and its stability function satisfies

$$\lim_{\substack{z\to-\infty \\ \Re(z)<0}} |L(z)| = 0. \tag{55}$$

For methods that are $A(\alpha)$-stable, if the condition (55) holds within the wedge $|\arg(-z)| < \alpha$, the method is called $L(\alpha)$-stable.

Using the stability functions derived in the previous sections, we analyze the L-stability properties of the three block hybrid methods.

Block method (9):

The stability function given by (49) is

$$L(z) = \frac{-z^2 - 52z - 150}{98z^3 - 197z^2 + 248z - 150}.$$

The numerator has degree 2 and the denominator has degree 3. As $z \to -\infty$ along the negative real axis, the leading term behavior yields

$$L(z) \sim \frac{-z^2}{98z^3} = -\frac{1}{98z} \to 0.$$

Since method (9) is $A(\alpha)$-stable with $\alpha = 89.06^\circ$ (as determined in Section 5), it follows that it is $L(\alpha)$-stable within the same wedge.

Block method (13):

The stability function in (52) is

$$L(z) = \frac{-3z^3 - 137z^2 - 585z - 768}{125z^3 - 503z^2 + 951z - 768}.$$

Here, the numerator and the denominator are both of degree 3. The limit as $z \to -\infty$ is the ratio of the leading coefficients:

$$\lim_{z \to -\infty} L(z) = \frac{-3}{125} = -0.024 \neq 0.$$

Although method (13) is A-stable (see Figure 2 and the analysis in Section 5), it does not satisfy condition (55). Hence, method (13) is not L-stable.

Block method (17):

The stability function in (54) is

$$L(z) = \frac{3z^3 + 59z^2 + 234z + 300}{216z^4 - 471z^3 + 707z^2 - 666z + 300}.$$

The degree of the numerator is 3 and that of the denominator is 4. As $z \to -\infty$,

$$L(z) \sim \frac{3z^3}{216z^4} = \frac{1}{72z} \to 0.$$

Given that method (17) is $A(\alpha)$-stable with $\alpha = 87.12°$, it is $L(\alpha)$-stable within this wedge.

This analysis demonstrates that the stability function not only facilitates the assessment of A-stability but also provides a straightforward means to check for L-stability by examining the asymptotic behavior as $z \to -\infty$. The stability function's rational form allows direct asymptotic analysis, making it preferable over stability polynomials for L-stability assessment, as the latter requires solving for root behavior as $z \to -\infty$ from $\pi(r, z) = 0$.

3.12 Experiments

In this subsection, three test problems are presented for experimental purposes. These problems are original constructions, not previously reported in the literature, and are designed to further examine the conditions for effectively applying stability polynomials and stability functions. Due to space constraints and to avoid repetition, the detailed derivation steps are omitted; however, the derivation procedure follows the one used to obtain block methods (9), (13), and (17). Interested readers may replicate the procedure to ascertain other properties of the test problems. The three experiments illustrate distinct stability characteristics and serve to validate the theoretical framework presented in the preceding sections.

3.12.1 Experiment 1

$$\begin{cases} f_{n+1} = \frac{984}{475\,h}\left(-y_{n+1} - \frac{227}{1312}y_n + \frac{1539}{1312}y_{n+\frac{4}{3}} - \frac{49}{984}hf_n - \frac{17}{492}hf_{n+2} + \frac{1}{328}hf_{n+3}\right), \\ f_{n+\frac{4}{3}} = -\frac{1200}{227\,h}\left(-y_{n+\frac{4}{3}} + y_{n+1} - \frac{23}{19440}\,hf_n + \frac{167}{1080}hf_{n+1} - \frac{4}{405}hf_{n+2} + \frac{29}{48600}hf_{n+3}\right), \\ y_{n+2} = -\frac{664}{227}y_{n+1} + \frac{891}{227}y_{n+\frac{4}{3}} + \frac{2}{227}hf_n - \frac{419}{681}hf_{n+1} + \frac{208}{681}hf_{n+2} - \frac{5}{681}hf_{n+3}\right), \\ y_{n+3} = \frac{875}{227}y_{n+1} - \frac{648}{227}y_{n+\frac{4}{3}} - \frac{25}{681}hf_n + \frac{800}{681}hf_{n+1} + \frac{1025}{681}hf_{n+2} + \frac{70}{227}hf_{n+3}\right). \end{cases} \tag{56}$$

3.12.2 Experiment 2

$$\begin{cases} y_{n+1} = y_n - \frac{9}{24}hf_n - \frac{19}{24}hf_{n+1} + \frac{5}{24}hf_{n+2} - \frac{1}{24}hf_{n+3}, \\ y_{n+2} = y_{n+1} - \frac{1}{24}hf_n + \frac{13}{24}hf_{n+1} + \frac{13}{24}hf_{n+2} - \frac{1}{24}hf_{n+3}, \\ y_{n+3} = y_{n+2} + \frac{8}{24}hf_{n+1} + \frac{32}{24}hf_{n+2} + \frac{8}{24}hf_{n+3}. \end{cases} \tag{57}$$

3.12.3 Experiment 3

$$\begin{cases} y_{n+1} = y_{n+2} + \frac{1}{1080}hf_n - \frac{7}{40}hf_{n+1} - \frac{7}{40}hf_{n+2} + \frac{53}{1080}hf_{n+3}, \\ y_{n+2} = y_n + \frac{41}{135}hf_n + \frac{8}{5}hf_{n+1} + \frac{3}{5}hf_{n+2} - \frac{4}{135}hf_{n+3}, \\ y_{n+3} = y_{n+1} - \frac{31}{1080}hf_n + \frac{17}{40}hf_{n+1} + \frac{57}{40}hf_{n+2} + \frac{329}{1080}hf_{n+3}. \end{cases} \tag{58}$$

3.13 Numerical simulation

In this subsection, the block hybrid schemes (9), (13), and (17) are implemented for the following problems, respectively:

3.13.1 Problem 1

$$\begin{cases} y_1' = -0.1y_1 - 199.9y_2, \quad y_1(0) = 5, \\ y_2' = -200y_2, \quad y_2(0) = 2. \end{cases} \tag{59}$$

3.13.2 Problem 2

$$\begin{cases} y_1' = -(2+\varepsilon^{-1})y_1(x) + \varepsilon^{-1}\left(y_2(x)\right)^2, \quad y_1(0) = 1, \\ y_2' = y_1(x) - y_2(x) - \left(y_2(x)\right)^2, \quad y_2(0) = 1. \end{cases} \tag{60}$$

3.13.3 Problem 3

$$\begin{cases} y_1' = 998y_1 + 1998y_2, \quad y_1(0) = 1, \\ y_2' = -999y_1 - 1999y_2, \quad y_2(0) = 0. \end{cases} \tag{61}$$

4 Implementation of the proposed block hybrid methods

This section provides a detailed description of how to implement the three derived block hybrid methods: The explicit method BHEM (9), the implicit method BHIM (13), and the implicit method BHIM (17). The implementation follows the standard block-by-block advancement strategy. For all three methods, the solution is advanced in blocks of fixed step size h. At each block step, values at r future points are computed simultaneously (for method (9), $r = 3$; for (13), $r = 4$; for (17), $r = 4$). Once the block unknowns are computed, the block window slides forward, and the process repeats.

4.1 Implementing the explicit method BHEM (9)

Method (9) is explicit, requiring no iteration. Its matrix form is

$$M^{(1)}Y_{n+1} = M^{(0)}Y_n + hB^{(1)}F_{n+1},$$

where $Y_{n+1} = [y_{n+1}, y_{n+2}, y_{n+3}]^T$ and $F_{n+1} = [f_{n+1}, f_{n+2}, f_{n+3}]^T$. Because $M^{(1)}$ is lower triangular with ones on the diagonal (see (19)), the solution can be obtained by forward substitution without inverting the matrix.

4.2 Implementing the implicit methods BHIM (13) and BHIM (17)

Implicit methods require solving a system of nonlinear equations at each block. For a system of ODEs $y' = f(t, y)$ with d components, the block unknowns $Y_{n+1} \in \mathbb{R}^{rd}$ are obtained from

$$(M^{(1)} - zB^{(1)})Y_{n+1} = (M^{(0)} + zB^{(0)})Y_n, \quad \text{for method (13)},$$

$$(M^{(1)} - zB^{(1)})Y_{n+1} = M^{(0)}Y_n, \quad \text{for method (17)},$$

We write the block methods as $G(Y_{n+1}) = 0$, where G is the residual. For methods (13) and (17), the residuals for the jth component at each block point are defined by (22) and (27). We assemble the rd-dimensional unknown vector $\mathbf{Y} = [y_{n+1,1}, \ldots, y_{n+1,d}, y_{n+2,1}, \ldots, y_{n+r,d}]^T$. The Jacobian matrix $J = \partial G/\partial \mathbf{Y}$ is formed using the Jacobian of f, denoted $J_f = \partial f/\partial y$. The iteration used is

$$J(\mathbf{Y}^{(m)})\,\Delta\mathbf{Y}^{(m)} = -G(\mathbf{Y}^{(m)}), \quad \mathbf{Y}^{(m+1)} = \mathbf{Y}^{(m)} + \Delta\mathbf{Y}^{(m)}.$$

The iteration stops when $\|\Delta\mathbf{Y}^{(m)}\|_\infty < \text{tol}$.

5 Results and discussion

The stability polynomial $\pi(r, z)$ of (36) focuses on the positions of the characteristic roots r in the complex plane. Figure 1 demonstrates that BHEM (9) is $A(\alpha)$-stable. In contrast, the stability function $L(z)$ of (49) examines the complex parameter z and the positions of its zeros and poles, as shown in Figure 4 for the same block method. Notably, the stability functions (49), (52), and (54) for methods (9), (13), and (17) involve only z, whereas the stability polynomials (36), (39), and (42) explicitly contain the roots r. Consequently, Figures 1, 2, and 3 display stability regions based on the roots r of the characteristic equations, while Figures 4, 5, and 6 illustrate stability properties through the complex variable z for the stability functions.

Figure 4 indicates that the stability function $L(z)$ for block method (9) satisfies condition (44) of Definition 3, being analytic for all $\Re(z) < 0$ because all poles of $L(z)$ in (49) lie in the right half-plane. However, the mere absence of left-half-plane poles does not guarantee A-stability. From (43) and (49), restated here as,

$$|L(iy)| \leqslant 1, \quad \text{for all real } y;$$

and

$$L(z) = \frac{-z^2 - 52\,z - 150}{98\,z^3 - 197\,z^2 + 248\,z - 150}.$$

A complex number $z = x + iy$ on the imaginary axis has $z = 0 + iy$, that is, $z = iy$. Taking $y = -0.9$ and applying condition (43) to (49) gives

$$|L(-0.9i)| = \left| \frac{-(-0.9i)^2 - 52\,(-0.9i) - 150}{98\,(-0.9i)^3 - 197\,(-0.9i)^2 + 248\,(-0.9i) - 150} \right|.$$

Using the evalf function of Maple 2022.0 for numerical evaluation gives

$$|L(-0.9i)| = 1.028270409890 > 1,$$

which violates condition (43) of Definition 3 because the value obtained at $y = -0.9$ is 1.028270409890, which is not less than or equal to 1, as required. However, it can be noted that the choice of $y = -0.9$ was based on choosing from the lower interval of the imaginary axis based on common patterns of stability plots. If there exists at least one y in such an interval for which (43) is not satisfied, then condition (43) is violated. This is because $|L(iy)|$ must be less than or equal to 1 for all real y. More comprehensively, applying Theorem 1 to method (9) shows that it fails both conditions for A-stability. First, on the imaginary axis, numerical investigation reveals that for $y = -0.9$, at least one root r of $\pi(r, iy) = 0$ satisfies $|r| > 1$. Second, in the left half-plane, the stability region shown in Figure 1 does not entirely cover $\{z : \Re(z) < 0\}$; there exist z with negative real part for which some root r of $\pi(r, z) = 0$ has $|r| \geq 1$. Therefore, method (9) is not A-stable. Nevertheless, verification confirms it is $A(\alpha)$-stable with $\alpha = 89.06°$, meaning absolute stability holds within the wedge $|\arg(-z)| < 89.06°$.

The stability polynomial $\pi(r, z)$ for block method (13), given by (39), is plotted in Figure 2. The region of absolute stability, corresponding to the exterior of the enclosed shape, encompasses the entire left half-plane, indicating that method (13) is A-stable. Correspondingly, the stability function $L(z)$ in (52) satisfies $|L(z)| \leqslant 1$ and is analytic for $\Re(z) < 0$, as shown in Figure 5. All poles of $L(z)$ lie in the right half-plane, and all zeros lie in the left half-plane. Since no pole occurs in the left half-plane, $L(z)$ is analytic for $\Re(z) < 0$, satisfying condition (44) of Definition 3. Moreover, $|L(iy)| \leqslant 1$ for all real y, fulfilling condition (43) of Definition 3. Applying Theorem 1 confirms A-stability: first, on the imaginary axis, all roots of $\pi(r, iy) = 0$ satisfy $|r| \leq 1$ for every real y; second, in the left half-plane, the stability region covers $\{z : \Re(z) < 0\}$ entirely, and for every such z, all roots of $\pi(r, z) = 0$ satisfy $|r| < 1$. Consequently, method (13) is A-stable.

The stability polynomial $\pi(r, z)$ for block method (17), given by (42), is plotted in Figure 3. The region of absolute stability (the exterior of the enclosed portion) covers part of the left half-plane but not all of it, indicating that method (17) is $A(\alpha)$-stable. Using (33) yields $\alpha = 87.12°$. The stability function $L(z)$ in (54), plotted in Figure 6, shows all poles in the right half-plane, so $L(z)$ is analytic for $\Re(z) < 0$, satisfying condition (44) of Definition 3. However, for $y = -0.98$, we compute $|L(iy)| = 1.204625225 > 1$, violating condition (43) of Definition 3. Equivalently, $E(y) = -34356.13668$ for (54), which contradicts condition (46). Applying Theorem 1 confirms that method (17) is not A-stable: first, on the imaginary axis, for $y = -0.98$ at least one root of

$\pi(r, iy) = 0$ has modulus exceeding unity; second, in the left half-plane, the stability region does not cover the entire set $\{z : \Re(z) < 0\}$. Hence, method (17) is not A-stable but is $A(\alpha)$-stable with $\alpha = 87.12°$, meaning absolute stability holds within the wedge $|\arg(-z)| < 87.12°$.

Simple algebraic simplification of Experiment 1 converts it to the form:

$$\begin{aligned}
y_{n+1} =& -\frac{227}{1312}y_n + \frac{1539}{1312}y_{n+\frac{4}{3}} - \frac{49}{984}hf_n - \frac{475}{984}hf_{n+1} - \frac{17}{492}hf_{n+2} \\
&+ \frac{1}{328}hf_{n+3}, \\
y_{n+\frac{4}{3}} =& y_{n+1} - \frac{23}{19440}hf_n + \frac{167}{1080}hf_{n+1} + \frac{227}{1200}hf_{n+\frac{4}{3}} - \frac{4}{405}hf_{n+2} \\
&+ \frac{29}{48600}hf_{n+3}, \\
y_{n+2} =& -\frac{664}{227}y_{n+1} + \frac{891}{227}y_{n+\frac{4}{3}} + \frac{2}{227}hf_n - \frac{419}{681}hf_{n+1} + \frac{208}{681}hf_{n+2} \\
&- \frac{5}{681}hf_{n+3}, \\
y_{n+3} =& \frac{875}{227}y_{n+1} - \frac{648}{227}y_{n+\frac{4}{3}} - \frac{25}{681}hf_n + \frac{800}{681}hf_{n+1} + \frac{1025}{681}hf_{n+2} \\
&+ \frac{70}{227}hf_{n+3}.
\end{aligned} \tag{62}$$

The stability polynomial obtained using (32) is

$$\begin{aligned}
\pi(r, z) =& -\frac{227\,r^4}{1312} + \frac{227\,r^3}{1312} + \frac{227\,r^3z^4}{15744} + \frac{2497\,r^4z}{9840} - \frac{4313\,r^4z^2}{26240} \\
&+ \frac{7037\,r^4z^3}{118080} - \frac{227\,r^4z^4}{19680} + \frac{9223\,r^3z}{31488} + \frac{4767\,r^3z^2}{26240} + \frac{46183\,r^3z^3}{472320}.
\end{aligned} \tag{63}$$

Applying the same procedure used to obtain (49), (52), and (54) yields the stability function for Experiment 1:

$$L(z) = \frac{6810\,z^4 + 46183\,z^3 + 85806\,z^2 + 138345\,z + 81720}{5448\,z^4 - 28148\,z^3 + 77634\,z^2 - 119856\,z + 81720}. \tag{64}$$

Plots of (63) and (64) are shown in Figure 7. Although no poles of the stability function $L(z)$ in (64) exist in the left half-plane, condition (43), or equivalently condition (46), is not satisfied. For instance, consider $y = -3$; following a similar approach to that used earlier, it can be shown that $|L(iy)| = 1.928808042740$ or $E(y) = -520155100224$, contradicting the requirements of (43) or (46). Thus, the block method in Experiment 1 is not A-stable because it violates condition (1) of Definition 4. While methods that are not A-stable may be explored for $A(\alpha)$-stability with sufficiently large α [29], stability functions $L(z)$ provide limited information on $A(\alpha)$-stability, whereas stability polynomials $\pi(r, z)$ offer more insight. The stability polynomial $\pi(r, z)$ in (63)

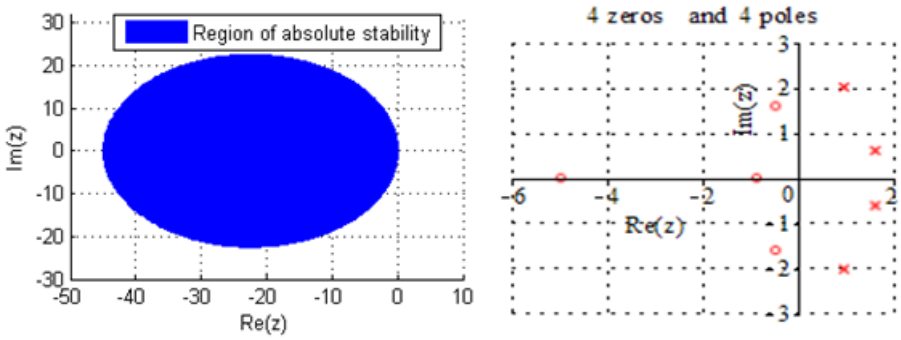


Figure 7: a) (Left) Plot of stability polynomial (63); b) (Right) Plot of stability function (64).

indicates that the block method of Experiment 1 has a very restricted region of absolute stability, as shown in its interval of absolute stability for the enclosed region on the left-hand side of the complex plane in Figure 7a. Placing wedges on the bounded region will not yield an infinite region on the left-hand side because it has a finite interval of absolute stability $(-45 < h\lambda < 0)$. Hence, the block method in Experiment 1 is neither A-stable nor $A(\alpha)$-stable but stable only in the interval $(-45 < h\lambda < 0)$. The block method in Experiment 1 does not possess an infinite interval of absolute stability like the block method (9), whose plot is shown in Figure 1; therefore, Experiment 1 should not be considered for stiff DEs. Using such methods for the solution of stiff DEs will, according to Dahlquist [9], impose prohibitive restrictions on the step size h.

For Experiment 2, the stability polynomial is obtained as

$$\pi(r,z) = r^3 - r^2 - \frac{5}{24}r^3z - \frac{3}{8}\,r^3z^2 - \frac{43}{24}r^2z + \frac{3}{8}\,r^2z^2 + \frac{1}{4}\,r^3z^3 + \frac{1}{4}\,r^2z^3 \tag{65}$$

and the stability function is

$$L(z) = \frac{-6\,z^3 - 9\,z^2 + 43\,z + 24}{6\,z^3 - 9\,z^2 - 5\,z + 24}. \tag{66}$$

Plots of (65) and (66) are shown in Figure 8.

Figure 8b shows that the stability function $L(z)$ in (66) has a pole on the left-hand side of the complex plane; hence, condition (44) of Definition 3 is not satisfied. Therefore, the block method in Experiment 2 cannot be A-stable [26]. It can also be verified that conditions (43) and (44) are not met. The plot of the stability polynomial $\pi(r,z)$ in Figure 8a indicates a more restricted region with an interval of absolute stability $(-0.9 < h\lambda < 0)$. Hence, the block method in Experiment 2 is also neither A-stable nor $A(\alpha)$-stable but stable only in the interval

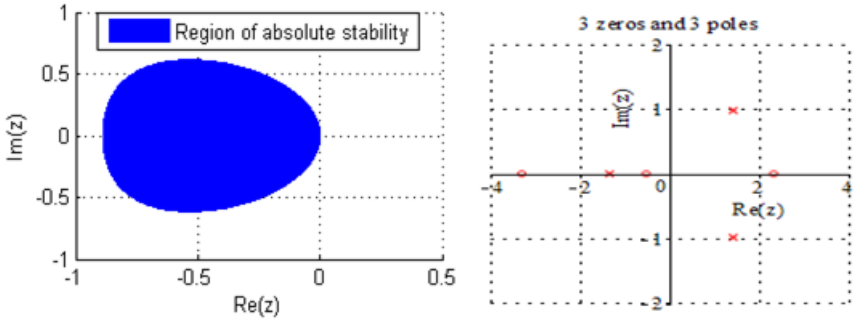


Figure 8: a) (Left) Plot of stability polynomial (65); b) (Right) Plot of stability function (66).

$(-0.9 < h\lambda < 0)$ and should not be considered for solving stiff DEs. However, it may be considered for nonstiff DEs, as discussed for Experiment 1.

For Experiment 3, the stability polynomial and stability function are

$$\begin{aligned}\pi(r,z) =& r^3 - r^2 - \frac{2537}{1080}r^3 z - \frac{2107}{1080}r^2 z + \frac{9181}{10800}r^3 z^2 - \frac{89}{120}r^2 z^2 \\ & - \frac{59}{400}r^3 z^3 - \frac{1}{20}\,r^2 z^3 \end{aligned} \tag{67}$$

and

$$L(z) = \frac{-540\,z^3 - 8010\,z^2 - 21070\,z - 10800}{1593\,z^3 - 9181\,z^2 + 25370\,z - 10800}. \tag{68}$$

Plots of (67) and (68) are shown in Figure 9.

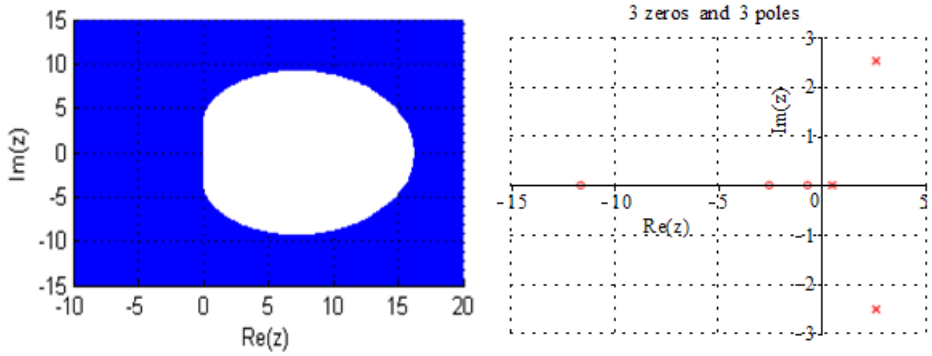


Figure 9: a) (Left) Plot of stability polynomial (67); b) (Right) Plot of stability function (68).

It can be verified that conditions (43) and (44) of Definition 3 are satisfied for the stability function $L(z)$ in (68), and the plot in Figure 9b shows no poles on the left-hand side and

$|L(z)| \leqslant 1$ for all $y \in \mathbb{R}$. Hence, the block method of Experiment 3 is A-stable. As confirmatory evidence, the plot of the stability polynomial $\pi(r, z)$ in (67) (Figure 9a) shows that the area outside the bounded region (blue) covers the entire left half-plane, implying an infinite interval $(-\infty < h\lambda < 0)$ of absolute stability. Thus, the block method in Experiment 3 is A-stable.

In summary, the stability function $L(z)$ can determine whether a numerical method is A-stable, but it may be limited in determining the angle α for $A(\alpha)$-stable methods. Nevertheless, stability functions provide a direct test for A-stability and are better suited for assessing L-stability properties because the stability function's rational form allows direct asymptotic analysis, making it preferable over stability polynomials for L-stability assessment, as the latter requires solving for root behavior as $z \to -\infty$ from $\pi(r, z) = 0$. Stability polynomials, however, can determine both A-stability and the angle α for $A(\alpha)$-stable methods. Additionally, methods that are neither A-stable nor $A(\alpha)$-stable can be shown to have a finite interval of absolute stability. A drawback of the stability polynomial $\pi(r, z)$ is that $A(\alpha)$-stable methods with very large α may have plots around the origin that appear A-stable. This can be addressed by adjusting the scale to see an amplified version of the plots, especially around the origin, as shown in Figures 1, 2, and 3. This enables clear visualization and valid conclusions. Stability polynomials and stability functions serve complementary theoretical roles. For consistent methods, the two approaches must yield congruent stability conclusions. A discrepancy between them serves as a theoretical indicator of potential derivation errors or method inconsistency.

Results from the numerical simulation of Problem 1 using BHEM (9) are shown in Figure 10.

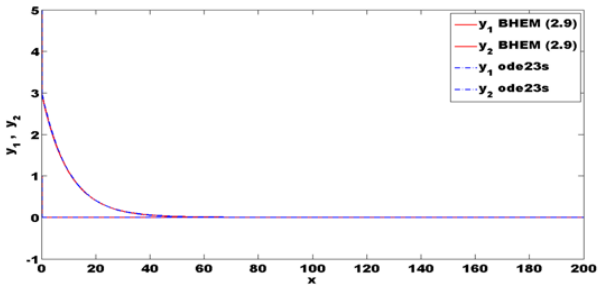


Figure 10: Numerical simulation of Problem 1 using BHEM (9).

Results from the numerical simulation of Problem 2 using the block hybrid implicit method BHIM (13) are shown in Figure 11.

The numerical simulation of Problem 3 using BHIM (17) yields the results shown in Figure 12.

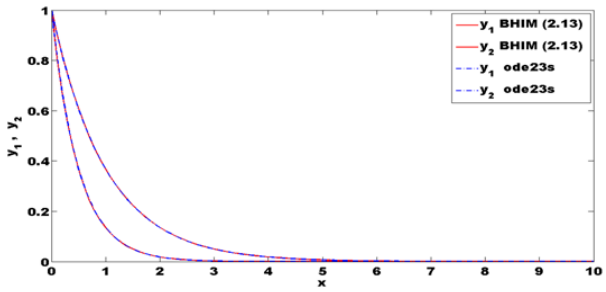


Figure 11: Numerical simulation of Problem 2 using BHIM (13).

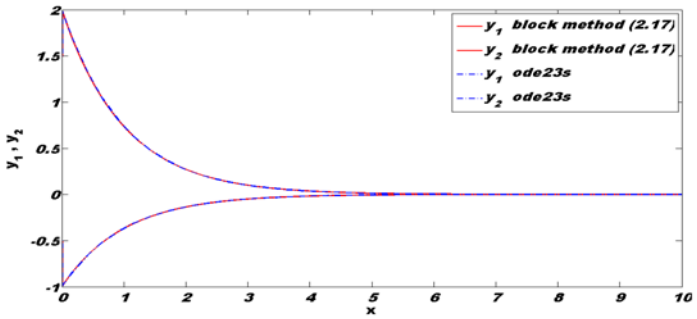


Figure 12: Numerical simulation of Problem 3 using BHIM (17).

Problem 1 is moderately stiff according to Fatunla [11], and Problem 2 is very stiff as $\varepsilon \to 0$. From Figures 10, 11, and 12, it can be seen that the block methods (9), (13), and (17) perform comparably to the built-in MATLAB `ode23s` solver for numerical solutions of Problems 1, 2, and 3.

To compare the accuracy of the results obtained from the block methods (9), (13), and (17) with exact solutions and the Adomian decomposition method reported in the literature, consider the following system:

$$\begin{cases} \dfrac{dx}{d\tau} = x + y, \quad x(0) = 0, \\ \dfrac{dy}{d\tau} = -x + y, \quad y(0) = 1, \end{cases} \tag{69}$$

whose exact solutions are given by

$$\begin{cases} x = e^{\tau} \sin \tau, \\ y = e^{\tau} \cos \tau. \end{cases}$$

The exact solutions from the system of ODEs (69) are compared with solutions from the block methods (9), (13), and (17) and the Adomian decomposition method described in the literature [27]. These comparisons are presented in Tables 1, 2, and 3.

Table 1: Comparison between the exact solutions (x and y), block method ($x_{eq(9)}$ and $y_{eq(9)}$) and Adomian decomposition method (x_ϕ and y_ϕ) solutions.

τ	Exact x	Exact y	$x_{eq(9)}$	$y_{eq(9)}$	x_ϕ [27]	y_ϕ [27]
0.50	0.79044	1.44689	0.79044	1.44689	0.79044	1.44688
1.00	2.28736	1.46869	2.28736	1.46869	2.28733	1.46868
1.50	4.47046	0.31702	4.47046	0.31702	4.47040	0.31701
2.00	6.71885	-3.07493	6.71885	-3.07493	6.71872	-3.07488
2.50	7.29088	-9.75993	7.29088	-9.75993	7.29069	-9.75970
3.00	2.83447	-19.88450	2.83447	-19.88453	2.83435	-19.88400
3.50	-11.61630	-31.01120	-11.61635	-31.01119	-11.61600	-31.01010
4.00	-41.32000	-35.68770	-41.32002	-35.68773	-41.31850	-35.68620
4.50	-87.99450	-18.97520	-87.99446	-18.97523	-87.99907	-18.97410

Table 2: Comparison between the exact solutions (x and y), block method ($x_{eq(13)}$ and $y_{eq(13)}$) and Adomian decomposition method (x_ϕ and y_ϕ) solutions.

τ	Exact x	Exact y	$x_{eq(13)}$	$y_{eq(13)}$	x_ϕ [27]	y_ϕ [27]
0.50	0.79044	1.44689	0.79043	1.44689	0.79044	1.44688
1.00	2.28736	1.46869	2.28733	1.46870	2.28733	1.46868
1.50	4.47046	0.31702	4.47042	0.31707	4.47040	0.31701
2.00	6.71885	-3.07493	6.71881	-3.07480	6.71872	-3.07488
2.50	7.29088	-9.75993	7.29095	-9.75965	7.29069	-9.75970
3.00	2.83447	-19.88450	2.83485	-19.88410	2.83435	-19.88400
3.50	-11.61630	-31.01120	-11.61531	-31.01082	-11.61600	-31.01010
4.00	-41.32000	-35.68770	-41.31797	-35.68805	-41.31850	-35.68620
4.50	-87.99450	-18.97520	-87.99141	-18.97758	-87.99907	-18.97410

Table 3: Comparison between the exact solutions (x and y), block method ($x_{eq(17)}$ and $y_{eq(17)}$) and Adomian decomposition method (x_ϕ and y_ϕ) solutions.

τ	Exact x	Exact y	$x_{eq(17)}$	$y_{eq(17)}$	x_ϕ [27]	y_ϕ [27]
0.50	0.79044	1.44689	0.79058	1.44690	0.79044	1.44688
1.00	2.28736	1.46869	2.28777	1.46851	2.28733	1.46868
1.50	4.47046	0.31702	4.47115	0.31613	4.47040	0.31701
2.00	6.71885	-3.07493	6.71924	-3.07738	6.71872	-3.07488
2.50	7.29088	-9.75993	7.28917	-9.76474	7.29069	-9.75970
3.00	2.83447	-19.88450	2.82694	-19.89126	2.83435	-19.88400
3.50	-11.61630	-31.01120	-11.63527	-31.01559	-11.61600	-31.01010
4.00	-41.32000	-35.68770	-41.35528	-35.67792	-41.31850	-35.68620
4.50	-87.99450	-18.97520	-88.04313	-18.92789	-87.99907	-18.97410

From the results in Tables 1, 2, and 3, it can be observed that the newly derived block methods (9), (13), and (17) perform competitively when compared with the exact solutions of system (69), and when compared and with the Adomian decomposition method.

5.1 CPU time performance of the proposed block methods

To assess the computational efficiency of the derived block hybrid methods, CPU time measurements were performed for Problems 1–3 described in Section 3.12. All codes were implemented in MATLAB R2018a on a system with an Intel Core i7-4900MQ processor, 32 GB RAM, and a 512 GB SSD, with a Newton iteration tolerance of 10^{-6}. The explicit method BHEM (9), applied to Problem 1, required 2.5788 seconds of CPU time, reflecting the higher cost associated with its stability constraints. In contrast, the implicit method BHIM (13) demonstrated remarkable efficiency for the moderately stiff Kaps Problem 2, completing the integration in just 0.1235 s of CPU time, a performance attributable to its A-stability, which permits larger step sizes. The implicit method BHIM (17), applied to the severely stiff linear system (Problem 3), required 2.3805 seconds of CPU time.

For system (69), CPU time measurements were performed for both short-range and long-range integration. The measured CPU times are summarized in Table 4.

Table 4: CPU times (in seconds) of the proposed block methods for (69).

Method	**Interval**	**CPU Time (s)**
BHEM (9)	$[0, 10]$	1.9762
BHIM (13)	$[0, 300]$	0.7544
BHIM (17)	$[0, 300]$	0.4779

The explicit method BHEM (9), which has weaker stability properties, was applied over $[0, 10]$ with $h = 0.001$, requiring 1.9762 s even over the short interval. In contrast, the implicit methods BHIM (13) and BHIM (17) were applied over the extended interval $[0, 300]$ with $h = 0.1$, requiring 0.7544 s and 0.4779 s, respectively. These results reveal a clear distinction in computational efficiency between the explicit and implicit formulations. The explicit method incurs a significantly higher cost due to the small step size necessary to maintain stability, making it suitable only for moderate integration ranges. The implicit methods, by contrast, leverage their superior stability properties—A-stability for BHIM (13) and $A(\alpha)$-stability with $L(\alpha)$-stability for BHIM (17)—to permit substantially larger step sizes, resulting in far lower CPU times despite integrating over a 30-fold longer interval. Notably, BHIM (17) is the most efficient among the proposed methods, completing the long-range integration in about half the time of BHIM (13). Overall, these findings confirm that the implicit block formulations, particularly BHIM (17), offer a compelling balance of accuracy and efficiency for long-time integration of problems with rapid exponential growth.

6 Conclusion

This study has compared the stability polynomial $\pi(r, z)$ and the stability function $L(z)$ for investigating the absolute stability regions of block hybrid methods. The analysis demonstrates that the stability polynomial focuses primarily on the characteristic roots r, while the stability function concentrates on the complex parameter z and the positions of its zeros and poles.

The research question posed in the introduction receives a clear answer: the absence of poles in the left half-plane is necessary but not sufficient for A-stability. Condition (1) of Definition 4 (boundedness on the imaginary axis) must also be satisfied, as illustrated by methods (9) and (17), which have no left-half-plane poles yet are not A-stable. This finding corrects a potential misconception in the literature that analyticity alone guarantees A-stability. The main contributions of the work are as follows: a novel theoretical result stating that for consistent block methods, the two stability tools must yield congruent stability conclusions, with any discrepancy serving as an indicator of potential derivation errors or method inconsistency; three practical rec-

ommendations, (i) when using stability polynomials, plots should be examined near the origin to accurately assess A-stability, (ii) when employing stability functions, both associated conditions must be verified, and (iii) a combined approach using both tools is advisable for comprehensive stability investigation; a demonstration that the stability function is preferable for analyzing stronger stability properties such as L-stability and $L(\alpha)$-stability, owing to its direct rational form that allows immediate asymptotic analysis; and a detailed stability analysis of three new block hybrid methods—one explicit (BHEM) and two implicit (BHIM)—using both stability polynomials and stability functions.

The use of stability polynomials offers a straightforward determination of the angle α for methods that are $A(\alpha)$-stable. A drawback is that an $A(\alpha)$-stable method with a large α may be mistakenly identified as A-stable if the plot is not examined at an appropriate scale near the origin.

For stability functions, when both prescribed conditions are satisfied, they provide a clear means to determine whether a block hybrid method is A-stable. Some researchers have used only one of these conditions to conclude A-stability; however, satisfying only one condition does not guarantee A-stability. Both conditions must be examined. A limitation of stability functions is their reduced effectiveness in determining the angle α for $A(\alpha)$-stable methods, as they focus primarily on the positions of zeros and poles. Nevertheless, stability functions are valuable for investigating stronger stability properties such as L-stability.

In light of these findings, we recommend that stability be examined using both stability polynomials and stability functions to ensure a comprehensive and reliable analysis. The two approaches are complementary: stability polynomials are particularly useful for identifying $A(\alpha)$-stability angles and finite stability intervals, while stability functions provide a direct test for A-stability and are better suited for assessing L-stability properties. For consistent methods, agreement between the two approaches validates the derivation, whereas discrepancies may signal errors or inconsistencies.

Acknowledgements

The contributions of Professor G.M. Kumleng of the Department of Mathematics, University of Jos, to the simulations of block hybrid methods are gratefully acknowledged.

Funding

The authors neither applied for nor received any external funding for this research.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Adedire, O., Sirisena, U.W. and Ndam, J.N., *Implementation of an A-stable block hybrid method for the solution of Bazykin ecological model and Trambouze reactor model*, New Trends Math. Sci., 8(2) (2020), 1–8.

[2] Akinfenwa, O.A., Abdulganiy, R.I., Akinnukawe, B.I. and Okunuga, S.A., *Seventh order hybrid block method for solution of first order stiff systems of initial value problems*, J. Egypt. Math. Soc., 28(34) (2020), 1–11, https://doi.org/10.1186/s42787-020-00095-3.

[3] Akinfenwa, O.A. and Jator, S.N., *Continuous block backward differentiation formula for solving stiff ordinary differential equations*, Comput. Math. Appl., 65 (2013), 996–1005, http://dx.doi.org/10.1016/j.camwa.2012.03.111.

[4] Akinfenwa, O.A., Jator, S.N. and Yao, N.M., *On the stability of continuous block backward differentiation formula for solving stiff ordinary differential equations*, J. Mod. Methods Numer. Methods, 3(2) (2012), 50–58, https://doi.org/10.1186/s42787-020-00095-3.

[5] Akinnukawe, B.I., Akinfenwa, O.A. and Okunuga, S.A., *L-stable block backward differentiation formula for parabolic partial differential equations*, Ain Shams Eng. J., 7 (2016), 867–972, http://dx.doi.org/10.1016/j.asej.2015.12.012.

[6] Celaya, E.A. and Anza, J.J., *BDF-α: A multistep method with numerical damping control*, J. Comput. Math., 1(3) (2013), 96–108, doi: 10.13189/ujcmj.2013.010305.

[7] Chu, M.T. and Hamilton, H., *Implicit parallel solution of ODEs by multi-block methods*, SIAM J. Sci. Stat. Comput., 8(3) (1987), 342–353.

[8] Dahlquist, G., *Convergence and stability in the numerical integration of ordinary differential equations*, Math. Scand., 4 (1956), 33–53, https://doi.org/10.7146/math.scand.a-10454.

[9] Dahlquist, G.G., *A special stability problem for linear multistep methods*, BIT Numer. Math., 3 (1963), 27–43, https://doi.org/10.1007/BF01963532.

[10] Fatunla, S.O., *Numerical methods for initial value problems in ordinary differential equations*, Academic Press, 1978.

[11] Fatunla, S.O., *A class of block methods for second order IVPs*, Int. J. Comput. Math., 55 (1995), 119–133, https://doi.org/10.1080/00207169508804368.

[12] Fatunla, S.O., *Block methods for second order ODEs*, Int. J. Comput. Math., 41 (1991), 55–63.

[13] Hairer, E., Lubich, C. and Wanner, G., *Geometric numerical integration: Structure-preserving algorithm for ordinary differential equations*, Springer Berlin, Heidelberg, 2006, https://doi.org/10.1007/3-540-30666-8.

[14] Hairer, E. and Wanner, G., *Solving ordinary differential equations II: stiff and differential algebraic problems (2nd ed.)*, Springer-verlag, Berlin, 1996, doi: 10.1007/978-3-642-05221-7.

[15] Henrici, P., *Discrete variable methods in ODEs*, John Wiley, 1962.

[16] Ibrahim, Z.B., Othman, K.I. and Suleiman, M., *Implicit r-point block backward differentiation formula for solving first-order stiff ODEs*, Appl. Math. Comput., 186 (2007), 558–565.

[17] Jator, S.N., Sahi, R., Akinyemi, M. and Nyonna, D., *Exponentially fitted block backward differentiation formulas for pricing options*, Cogent Econ. Finance, 9 (2021), 1875565, doi: 10.1080/23322039.2021.1875565.

[18] Ngwane, F.F. and Jator, S.N., *L-Stable block hybrid second derivative algorithm for parabolic partial differential equations*, Am. J. Comput. Math., 4 (2014), 87–92, http://dx.doi.org/10.4236/jasmi.2014.42008.

[19] Okuonghae, R.I. and Ikhile, M.N.O., *A class of hybrid linear multistep methods with $A(\alpha)$-stability properties for stiff IVPs in ODEs*, J. Numer. Math., 21(2) (2013), 157–172, doi: 10.1515/jnum-2013-0006.

[20] Okuonghae, R.I. and Ikhile, M.N.O., *A family of highly stable second derivative block methods for stiff IVPs in ODEs*, Numer. Anal. Appl., 7(1) (2014), 57–69, https://doi.org/10.1134/S1995423914010066.

[21] Okuonghae, R.I. and Ikhile, M.N.O., *Stiffly stable second derivative linear multistep methods with two hybrid points*, Numer. Anal. Appl., 8 (2015), 248–259, https://doi.org/10.1134/S1995423915030052.

[22] Onumanyi, P., Sirisena, U. W. and Jator, S. N., *Continuous finite difference approximations for solving differential equations*, Int. J. Comput. Math., 72(1) (1999), 15–27, doi: 10.1080/00207169908804831.

[23] Oyelami, O.A., Gideon, I.E., Mohammed, U., Akeem, D., Garba, J. and Elebor, C.C., *Optimization of two hybrid points in a one step method*, J. NAMP, 72 (2026), 11–22, https://doi.org/10.60787/jnamp.vol72no.658.

[24] Ramos, H. and Rufai, M., *A two-step hybrid block method with fourth derivatives for solving third-order boundary value problems*, J. Comput. Appl. Math., 404 (2021), 113419, doi: 10.1016/j.cam.2021.113419.

[25] Ramos, H. and Rufai, M.A., *A new one-step method with three intermediate points in a variable step-size mode for stiff differential systems*, J. Math. Chem., 61 (2023), 673–688, https://doi.org/10.1007/s10910-022-01427-7.

[26] Shateyi, S., *On the application of the block hybrid methods to solve linear and nonlinear first order differential equations*, Axioms, 12(2) (2023), 158–189, https://doi.org/10.3390/axioms12020158.

[27] Shawagfeh, N. and Kaya, D., *Comparing numerical methods for the solutions of systems of ordinary differential equations*, Appl. Math. Lett., 17(2) (2004), 323–328.

[28] Tassaddiq, A., Qureshi, S., Soomro, A., Hincal, E. and Shaikh, A.A., *A new continuous hybrid block method with one optimal intrastep point through interpolation and collocation*, Fixed Point Theory Algorithms Sci. Eng., 22 (2022), 1–17, https://doi.org/10.1186/s13663-022-00733-8.

[29] Widlund, O.B., A note on unconditionally stable linear multistep methods, BIT, 7 (1967), 65–70.

Iranian Journal of Numerical Analysis and Optimization
Vol. 17, No. 1, 2027, pp 457–493
https://doi.org/10.22067/ijnao.2026.96965.1802
https://ijnao.um.ac.ir/

CC BY NC

Research Article

An approach to find the optimal design of a 3D damping set for a damped wave equation in spherical coordinates

H. Alimorad, A. Fakharzadeh Jahromi* and M. Goodarzi

Abstract

Damped waves play an important role in various aspects of modern life, such as analyzing ocean waves, storage tanks, automotive suspension systems, designing earthquake damping systems, and MRI machines in medicine. Furthermore, analyzing some of these systems is more naturally and efficiently performed in spherical coordinates. Considering these two important points, in this paper, we model the problem of finding the optimal shape of the damping set to minimize the energy of damped waves within a given end time defined in spherical coordinates. The basic idea is to describe the unknown 3D shape using two positive measures in spherical coordinates. This helps us to change the original problem into a linear optimization task that can be solved easily and globally. After explaining and modeling the problem and determining the weak

*Corresponding author

Received 13 December 2025; revised 16 July 2026; accepted 15 September 2026

Hajar Alimorad
Department of Mathematics, Jahrom University, P. O. Box: 74135-111, Jahrom, Iran. e-mail: h.alimorad@jahromu.ac.ir

Alireza Fakharzadeh Jahromi
Department of Mathematics, Shiraz University of Technology, Shiraz, Iran. e-mail: a_fakharzadeh@sutech.ac.ir

Mina Goodarzi
Department of Mathematics, Shiraz University of Technology, Shiraz, Iran. e-mail: M.godarzi01@yahoo.com

How to cite this article
Alimorad, H., Fakharzadeh Jahromi, A. and Goodarzi, M., An approach to find the optimal design of a 3D damping set for a damped wave equation in spherical coordinates. *Iran. J. Numer. Anal. Optim.*, 2027; 17(1): 457-493. https://doi.org/10.22067/ijnao.2026.96965.1802

solution in spherical coordinates, the measure-theoretical embedding process is also extended for the problem in spherical coordinates. First, for a fixed region, the problem is formulated as an optimal control in variational form and then converted into an infinite-dimensional linear optimization problem. To determine the nearly optimal shape of the damping set, we use an appropriate finite linear programming problem through an iterative method based on the particle swarm optimization algorithm. The advantages of this method over other methods are that it provides a linear treatment for our nonlinear optimization problem, has a global minimizer, and is independent of the initial shape and mesh design. To demonstrate the effectiveness of this method, some illustrative numerical examples are also provided.

AMS subject classifications (2020): 74J99; 90C90; 58Z05.

Keywords: Damping set; Measure theory; Optimal shape; PSO algorithm; Wave equation.

1 Introduction and background

In engineering, determining the best position and shape of sensors and actuators that minimize the energy of the system over a given time is a crucial concept [17]. Damped waves exhibit unique characteristics in three-dimensional (3D) space, including simultaneous propagation in all spatial directions, more complex wave interference in different directions, and more sophisticated calculations compared to two-dimensional (2D) cases, which lead to broader applications in realistic simulations. However, there are only a few classical methods for solving optimal shape design (OSD) problems in three dimensions, and they are generally less effective since, in most cases, they are prone to convergence issues. In this regard, the shape of the damping set has the greatest effect on the optimal rate of wave damping.

There are many studies on this optimization problem in the 2D domain. For instance, Münch, Pedregal, and Periago [25] determined the optimal damping set whose damping coefficient was fixed at a given value by applying the first-order gradient descent method. Then, the optimal damping set was determined in the case where the damping coefficient is also unknown by using the level set method [24]. Fakharzadeh, Alimorad, and Beiranvand found the optimal shape of a damping set and an unknown damping parameter simultaneously by using a measure-theoretical method [10].

It is worth noting that alternative methodologies for shape optimization of damped systems have also been explored. For instance, Zuazua [36] investigated geometric control conditions for wave energy decay, while Allaire [4] developed variational approaches for PDE-constrained shape optimization, and Henrot and Pierre [15] developed topological approaches (e.g., homogenization

and Hadamard derivatives) for PDE-constrained shape optimization. These methods, although powerful, often rely on smooth deformations or initial guesses, and may struggle with nonconvex or discontinuous design variables such as characteristic functions of damping sets.

To overcome the limitations of the classical methods mentioned above, measure-theoretical approaches have been developed. In 1999, an embedding method based on Radon measures for solving OSD problems in a 2D domain was established [13]. Other researchers have extended this method successfully for solving OSD problems in two dimensions [10, 14]; they called this method the shape measure. In this context, Mehne and Farahi [22] applied this measure-theoretic framework to transfer the variable domain of an OSD problem into a fixed reference domain, embedding the resulting problem into an infinite-dimensional linear programming (LP) model that was subsequently approximated by a finite-dimensional LP. It is worth noting that a complete overview of studies on optimal shape design using measures is listed in [8]. According to the above-mentioned studies, measure-theoretical methods are flexible, do not require any initial guess, and are not restricted by differentiability, convexity, or linearity of the cost or constraint functions.

Extending such approaches to three dimensions remains challenging. The first attempt to address a 3D symmetric optimal shape in cylindrical coordinates was made by Alimorad and Fakharzadeh [2]. Later, the measure-theoretical method was successfully extended to 3D cylindrical coordinates to solve obstacle problems [1], and to design optimal rotated shapes [7]. Although these studies demonstrated that the measure approach works well in 3D cylindrical domains, many real-world physical problems, such as damped waves in spherical tanks or acoustic fields, are naturally described in spherical coordinates. Solving these problems in cylindrical coordinates requires more complex mathematical formulations and introduces artificial boundaries. Therefore, using the extended measure-theoretical framework in spherical coordinates is necessary for solving such 3D physical problems accurately and easily. In this context, Fakharzadeh and Goodarzi [9] also adapted the shape-measure framework to spherical coordinates for designing nanoparticles, highlighting the growing need for spherical models.

While their approach successfully addressed this specific application, the present work introduces distinct and significant innovations compared to [9]. First, the physical and mathematical nature of the problems differs fundamentally. We recall that nanoparticle design typically involves elliptic or steady-state equations, whereas the damped wave equation is hyperbolic, characterized by finite propagation speeds, wave reflections, and complex interference patterns. Consequently, the energy functional involves both kinetic and potential energies, requiring weak solutions in different Sobolev spaces ($H^1 \times L^2$). Second, the objective of minimizing the total mechanical energy at a strict terminal time for a wave system introduces unique challenges in exact controllability and observability that are absent in steady-state designs. Therefore, the primary

motivation of this work is to bridge the gap between measure-theoretical OSD and complex 3D hyperbolic dynamics in spherical geometries.

Recently, several advanced methods have been developed for 3D shape optimization, which can be grouped into a few main categories. For isogeometric boundary element methods (IGBEM), researchers have combined IGBEM with adjoint variables to optimize 3D acoustic shapes [5, 30]. This method directly connects computer-aided design and numerical analysis, but still relies on gradient-based optimization.

A parametric level set approach was employed to design 3D thermoelastic metamaterials [33]. Furthermore, the level set method was improved by adding local volume constraints to control the shape during optimization [18]. However, these methods require a given initial shape and often fall into local minima.

Adjoint-based methods have been widely utilized to optimize 3D fins in heat exchangers and solar receivers [34, 35], as well as underwater body shapes [21]. Furthermore, this approach has been combined with volume penalization for compressible flows [20]. While highly effective, such techniques require smooth and differentiable objective functions, limiting their application to problems with discontinuous design variables.

Although these methods are powerful, they share four common limitations:

- They rely on gradients and can fall into local minima,
- They require a given initial shape to start,
- They require mesh updates or re-initialization when the shape changes,
- They cannot easily handle discontinuous design variables, such as the characteristic function of a damping set.

Recently, heuristic, meta-heuristic, and intelligent methods such as genetic algorithms (GA), particle swarm optimization (PSO), and artificial neural networks (ANNs) have been widely used to solve complex OSD and optimal control problems [11]. These methods are powerful tools for searching in large solution spaces. However, applying them directly to PDE-constrained shape optimization requires high computational costs and complex mesh updates. In this paper, we combine two methods. First, the measure-theoretical method transforms the complex, nonlinear shape optimization problem into a simple linear programming problem (LPP). Then, the PSO algorithm is used to find the 2D projection of the unknown 3D surface. This combination is very efficient and saves a lot of computing time compared to using only intelligent methods.

To overcome the mentioned challenges, the present work uses a measure-theoretical embedding scheme. This approach offers several distinct advantages:

- No initial shape required: By using positive Radon measures, the method finds the optimal shape without needing an initial geometric guess.

- Avoiding local minima traps: By transferring into a linear programming problem, the method avoids the local minima traps of gradient-based methods.

- Direct 3D formulation: The method works directly in full 3D spherical coordinates without needing artificial symmetry assumptions.

- Mesh independence: Unlike finite element or finite volume methods, this approach does not require mesh deformation and naturally manages discontinuous control variables.

From an optimal control perspective, although measure theory is a well-known mathematical tool, its application to this specific problem introduces a new and simpler computational strategy. In standard shape optimization for wave equations, the control variable (which defines the damping set) is discontinuous, which makes the optimization problem highly nonlinear and very hard to solve. The innovation of our measure-based method has two main parts:

- It handles the discontinuous damping set naturally. We do not need to artificially make the shape smooth or use very fine meshes near the boundaries.

- By using two positive measures, the method turns the complex, nonlinear shape optimization problem into a simple linear one that can be solved globally. This is a major step forward for solving 3D shape design problems for wave equations.

Despite these recent advances in 3D shape optimization, the problem of designing an optimal damping set for damped wave equations in spherical coordinates has not yet been studied. Most existing works focus on continuous boundaries or use coordinate systems that are not suitable for spherical shapes. This important gap motivates us to apply the measure-theoretical method specifically to this problem.

Based on these advantages, this paper focuses on finding the optimal shape of this damping set for the linear damped wave equation. The goal is to minimize the total mechanical energy at a given terminal time. This optimization problem is highly nonconvex and includes a discontinuous control (the characteristic function of the damping region). By using the flexibility of this method, we avoid mesh dependency, require no initial shape, and utilize powerful linearity properties.

To solve this problem, we use Green's formula to derive the weak solution in Sobolev spaces. This variational form naturally incorporates the boundary and initial conditions.

To determine the desired 3D optimal shape of the damping set, the organization of the paper is as follows: Section 2 presents the OSD problem, Section 3 presents its variational formulation, and Section 4 reformulates the problem in spherical coordinates. Section 5 transforms the problem into a classical optimal control problem, while Section 6 presents its measure-space formulation. Section 7 discusses the approximation schemes and methods used to determine the optimal shape, and section 8 presents the numerical examples.

2 Classical optimal shape design problem

Let $T > 0$ and let $\Sigma \subset \mathbb{R}^3$ be an open and bounded domain of class C^2 with a smooth boundary $\partial\Sigma$. Additionally, let σ be a measurable subset of Σ, which is independent of time $t \in (0, T)$. We consider the following linear damped wave equation in $(0, T) \times \Sigma$ with Dirichlet boundary conditions and initial data (u_0, u_1) [31]:

$$\begin{aligned} &\ddot{u} - \Delta u + a(x)\dot{u} = 0, && (t, x) \in (0, T) \times \Sigma, \\ &u = 0, && (t, x) \in (0, T) \times \partial\Sigma, \\ &u(0, x) = u_0(x), \quad \dot{u}(0, x) = u_1(x), && x \in \Sigma, \end{aligned} \tag{1}$$

where $\dot{u} = \frac{\partial u}{\partial t}$, $\ddot{u} = \frac{\partial^2 u}{\partial t^2}$ and Δ is the Laplacian with respect to the variable x. We assume that the damping potential $a(x) \in L^\infty(\Sigma, \mathbb{R}^+)$ has the following form [24]:

$$a(x) = a\chi_\sigma(x), \quad \text{for all } x \in \Sigma, \tag{2}$$

where $a \in \mathbb{R}^+$ and χ_σ is the characteristic function of σ. Indeed, $a(x)$ represents the damping potential in Σ and relation (2) ensures that the damping term $a(x)\,\dot{u}$ is active on σ. Moreover, the initial data (u_0, u_1) denote the initial position and velocity of the membrane whose displacement is governed by system (1) and are independent of both σ and a [25]. Moreover, $u(t, x)$ represents the transverse displacement of the membrane at point x at time t. For $(u_0, u_1) \in H_0^1(\Sigma) \times L^2(\Sigma)$, system (1) is well-posed [31]. In fact, there exists a unique weak solution of (1) in the class $C((0, T); H_0^1(\Sigma)) \cap C^1((0, T); L^2(\Sigma))$. According to [29], $H^1(\Sigma)$ denotes the usual Sobolev space, while $H_0^1(\Sigma)$ denotes the Sobolev space of order one on Σ whose elements vanish on the boundary of Σ.

Remark 1. A Sobolev space is a vector space of functions equipped with a norm that is a combination of L^p-norms of the function together with its derivatives up to a given order.

The energy at time t associated with system (1) has the following form [31]:

$$E(\sigma, a, t) = \frac{1}{2} \int_{\Sigma} \left(|\dot{u}|^2 + |\nabla u|^2 \right) dx. \tag{3}$$

Multiplying the first equation in (1) by $\dot{u}$, integrating over Σ and using integration by parts, we have from [26]

$$\dot{E}(\sigma, a, t) = -\int_{\Sigma} a(x) \left| \dot{u} \right|^2 dx, \quad \text{for all } t > 0.$$

Based on (2) and considering that $a \in \mathbb{R}^+$, $a(x) \geq 0$ and since $|\dot{u}|^2 \geq 0$, $\dot{E}(\sigma, a, t) \leq 0$ and hence the energy is a nonincreasing function of time.

We consider the OSD problem which consists of determining the best position and shape of σ in Σ that minimizes the energy of system (1) at a given end time T. Without restrictions on σ, the trivial solution is $\sigma = \Sigma$. To restrict the volume of σ, we introduce the following constraint [24]:

$$\Sigma_L = \{\sigma \subset \Sigma,\ volume\,(\sigma) = L \times\ volume\,(\Sigma)\}, \qquad 0 < L < 1. \tag{4}$$

Therefore, the mentioned OSD problem has the following form:

$$\min_{\sigma \in \Sigma_L} \ E(\sigma, a, T) = \frac{1}{2} \int_{\Sigma} (|\dot{u}|^2 + |\nabla u|^2)\, dx, \tag{5}$$

where u is the solution of system (1). The damping potential a may either be treated as a fixed parameter or considered an unknown variable, while optimizing σ and the aim here is to solve problem (5) using a two-phase optimization approach based on measure theory. For this purpose, first, the problem is transformed into a classical optimal control one in variational form. Next, it is presented in measure space as an infinite-dimensional optimization problem by defining two unknown positive Radon measures.

Although a similar problem was solved using cylindrical coordinates in [2], we prefer to seek the solution in spherical coordinates due to their specific advantages: Natural alignment with smooth boundaries, efficient representation of bounded domains, simple expression of Dirichlet boundary conditions, and convenient formulation of initial value problems–features that are particularly well suited to engineering and physical applications.

At the end of this section, we present three important and fundamental theorems that play a vital role in understanding the concepts introduced in subsequent sections.

Theorem 1 (Divergence theorem)**.** Suppose that Ω is a subset of $\mathbb{R}^n$ that is compact and has a piecewise smooth boundary Γ (indicated by $\partial\Omega = \Gamma$). If F is a continuously differentiable vector field defined on a neighborhood of Ω, then

$$\int_{\Omega} (\nabla \cdot \mathbf{F})\, d\Omega = \oint_{\Gamma} (\mathbf{F} \cdot \widehat{\mathbf{n}})\, \mathbf{d}\,,$$

where the left side is a volume integral over Ω, and the right side is the surface integral over the boundary of Ω [32].

Theorem 2 (Stokes' theorem)**.** Let S be a piecewise smooth, orientable surface bounded by a piecewise smooth closed curve C, and let $\mathbf{n}$ denote the unit normal vector to S. The orientations of S and C are assumed to be consistent according to the right-hand rule. In particular, when the fingers of the right hand follow the direction of traversal of C, the thumb points in the direction of $\mathbf{n}$. For a surface forming part of the boundary of a solid region, $\mathbf{n}$ points outward, and C is traversed in the positive (counterclockwise) direction when viewed from the side of the normal. The vector differential element of arc length along C is denoted by $\mathbf{dl}$.

Let $\mathbf{F}$ be a continuously differentiable vector field defined on an open region containing S. Then

$$\int_{S} \operatorname{curl} \mathbf{F} \cdot \mathbf{dA} = \oint_{C} \mathbf{F} \cdot \mathbf{dl}.$$

The oriented boundary of S is sometimes denoted by ∂S. Green's theorem can be viewed as a special case of Stokes' theorem applied to a planar surface [32].

Theorem 3 (Green's theorem in the plane)**.** Let D be a region in the xy-plane bounded by a piecewise smooth closed curve C. If $f(x, y)$ and $g(x, y)$ have continuous first partial derivatives on an open region containing D, then

$$\int_{D} \left(\frac{\partial g}{\partial x}(x, y) - \frac{\partial f}{\partial y}(x, y) \right) dA = \oint_{C} f(x, y)\, dx + g(x, y)\, dy,$$

where C is oriented counterclockwise.

For a vector field $\mathbf{F} = \langle F_1, F_2, F_3 \rangle$, when applying Stokes' theorem to a flat surface oriented upward in the xy-plane, we can identify $F_1(x, y, 0)$ with $f(x, y)$ and $F_2(x, y, 0)$ with $g(x, y)$. Since the surface lies in the xy-plane, F_3 vanishes identically. Furthermore, $\mathbf{dl}$ is perpendicular to $\mathbf{k}$ (the unit normal vector pointing in the positive z-direction), which explains why the third component does not contribute to the integral [32].

3 Weak solution

Establishing the existence of a classical solution u for problem (5) subject to system (1) is difficult. Therefore, weak (generalized) solutions are usually considered [23, 27]. In general, the main steps

presented in this paper could be introduced in this way. First, we state the classical problem as integral equations. Then, by introducing positive Radon measures, we change the problem into a measure space. In the third step, to overcome difficulties such as the existence of solutions, we extend the underlying space of the problem. In this step, we have an infinite linear programming problem in measure spaces. Finally, after solving the related LPP, the path functions and optimal control are calculated. This method has several advantages, such as the automatic existence of a solution, dealing with an alternative linear problem despite the nonlinearity of the original one, and the ability to find a global optimal solution for the problem. In this paper, this method has been extended for finding the optimal shape of damping set in 3D spherical coordinates to obtain the desired 3D unknown surface directly.

To solve (5) subject to system (1), first, the problem must be transformed into its variational form. For this purpose, we consider the following kind of Green's formula [32]:

$$\int_{\Sigma}(u\Delta V - V\Delta u)dx = \int_{\partial\Sigma}(u\frac{\partial V}{\partial n} - V\frac{\partial u}{\partial n})ds,$$

where u and V are both twice continuously differentiable on Σ and n is the outward-pointing unit normal vector of the surface. Let $V \in H_0^1(\Sigma \times (0,T))$; then, it is zero on the boundary $\partial\Sigma$ and $\int_{\partial\Sigma}(V\frac{\partial u}{\partial n})ds = 0$. Additionally, from equation system (1), we have $u = 0$ on the boundary $\partial\Sigma$ and $\int_{\partial\Sigma}(u\frac{\partial V}{\partial n})ds = 0$. As a result $\int_{\Sigma}(u\Delta V - V\Delta u)dx = 0$, and we obtain

$$\int_{\Sigma} V\Delta u\, dx = \int_{\Sigma} u\Delta V\, dx. \tag{6}$$

Similarly, using the following consequence of Green's formula:

$$\int_{\Sigma}(u\Delta V + \nabla u.\nabla V)dx = \int_{\partial\Sigma} u\,\nabla V.\, n\, ds,$$

Since $V \in H_0^1(\Sigma \times (0,T))$ is zero on the boundary $\partial\Sigma$ and by system (1), $u = 0$ on the boundary $\partial\Sigma$, then, $\int_{\partial\Sigma} u\,\nabla V.\, n\, ds = 0$ and hence we have

$$\int_{\Sigma} u\Delta V\, dx = -\int_{\Sigma} \nabla u\nabla V\, dx. \tag{7}$$

By (6) and (7), the following relation is obtained:

$$\int_{\Sigma} V\Delta u\, dx = -\int_{\Sigma} \nabla u\nabla V\, dx. \tag{8}$$

By multiplying the first equation in (1) with a function $V \in H_0^1(\Sigma \times (0,T))$ and applying (8), we have

$$\int_\Sigma \ddot{u}V\, dx + \int_\Sigma \nabla u.\nabla V\, dx + \int_\Sigma a(x)\dot{u}V\, dx = 0. \tag{9}$$

Integrating both sides of (9) with respect to t over $[0,T]$ gives

$$\int_0^T \int_\Sigma \ddot{u}V\, dxdt + \int_0^T \int_\Sigma \nabla u \nabla V\, dxdt + \int_0^T \int_\Sigma a(x)\dot{u}V\, dxdt = 0. \tag{10}$$

Double integration by parts with respect to t from the first term on the left-hand side of (10) shows that

$$\begin{aligned}\int_0^T \int_\Sigma \ddot{u}V\, dxdt = &\int_\Sigma (\dot{u}(T)V(T) - \dot{u}(0)V(0) - u(T)\dot{V}(T) \\ &+ u(0)\dot{V}(0))dx + \int_0^T \int_\Sigma u\ddot{V}\, dxdt.\end{aligned} \tag{11}$$

Similarly, integrating the third expression on the left-hand side of (10) gives

$$\begin{aligned}\int_0^T \int_\Sigma a(x)\dot{u}V\, dxdt = &\int_\Sigma a(x)(u(T)V(T) \\ &- u(0)V(0))\, dx - \int_0^T \int_\Sigma a(x)u\dot{V}\, dxdt.\end{aligned} \tag{12}$$

Now, the initial conditions of system (1) can be substituted into (11) and (12), respectively, to achieve

$$\begin{aligned}\int_0^T \int_\Sigma \ddot{u}V\, dxdt = &\int_\Sigma (\dot{u}(T)V(T) - u_1(x)V(0) \\ &- u(T)\dot{V}(T) + u_0(x)\dot{V}(0))dx + \int_0^T \int_\Sigma u\ddot{V}\, dxdt,\end{aligned} \tag{13}$$

$$\begin{aligned}\int_0^T \int_\Sigma a(x)\dot{u}V\, dxdt = &\int_\Sigma a(x)(u(T)V(T) \\ &- u_0(x)V(0))\, dx - \int_0^T \int_\Sigma a(x)u\dot{V}\, dxdt.\end{aligned} \tag{14}$$

By applying (13) and (14), equation (10) is transformed as follows:

$$\int_\Sigma \dot{u}(T)V(T)\,dx - \int_\Sigma u(T)\dot{V}(T)\,dx + \int_0^T \int_\Sigma u\ddot{V}\,dxdt$$
$$+ \int_0^T \int_\Sigma \nabla u.\nabla V\,dxdt + \int_\Sigma a(x)u(T)V(T)\,dx$$
$$- \int_\Sigma a(x)u_0(x)V(0)\,dx - \int_0^T \int_\Sigma a(x)u\dot{V}\,dxdt$$
$$= \int_\Sigma (u_1(x)V(0) - u_0(x)\dot{V}(0))\,dx. \tag{15}$$

The function u which satisfies (15) for each $V \in H_0^1(\Sigma \times (0,T))$ is called a weak solution of (1). Indeed, Green's formulas allow us to replace (1) with (15), which is simpler for the shape-measure method. Therefore, we change the problem and consider a new one with a different formulation. Additionally, constraint (4) on the volume of σ, can be shown by the following relation:

$$\int_\sigma dx = L\int_\Sigma dx, \qquad 0 < L < 1. \tag{16}$$

Based on Green's theorem, the variational formulation of problem (5) subject to system (1) is revised and can be written in the following form:

$$\min E(\sigma, a, t) = \frac{1}{2}\int_\Sigma \left(|\dot{u}|^2 + |\nabla u|^2\right) dx$$
$$subject\ to: \quad \int_\Sigma \dot{u}(T)V(T)\,dx - \int_\Sigma u(T)\dot{V}(T)\,dx + \int_0^T \int_\Sigma u\ddot{V}\,dxdt$$
$$+ \int_0^T \int_\Sigma \nabla u.\nabla V\,dxdt + \int_\Sigma a(x)u(T)V(T)\,dx$$
$$- \int_\Sigma a(x)u_0(x)V(0)\,dx - \int_0^T \int_\Sigma a(x)u\dot{V}\,dxdt$$
$$= \int_\Sigma (u_1(x)V(0) - u_0(x)\dot{V}(0))\,dx,$$
$$\int_\sigma dx = L\int_\Sigma dx, \qquad 0 < L < 1. \tag{17}$$

4 Expressing the problem in a spherical coordinate system

Based on the abilities of spherical coordinates for simplifying the formulation of some special shapes, we prefer to express and solve problem (5) in spherical coordinates. In this regard, for each $u(t,\rho,\varphi,\theta) \in C^2(\mathbb{R}^4)$, we have from [6]

$$\nabla u\,|_{_T} = \left(\frac{\partial u}{\partial \rho}, \frac{1}{\rho}\frac{\partial u}{\partial \varphi}, \frac{1}{\rho \sin\varphi}\frac{\partial u}{\partial \theta}\right).$$

Therefore, (3) can be expressed as follows:

$$E(\sigma, a, T) = \frac{1}{2}\int_{\Sigma} (|\dot{u}(\rho,\varphi,\theta,T)|^2 + (\frac{\partial u}{\partial \rho})^2 + \frac{1}{\rho^2}(\frac{\partial u}{\partial \varphi})^2 + \frac{1}{\rho^2 \sin^2\varphi}(\frac{\partial u}{\partial \theta})^2)\rho^2 \sin\varphi\, d\rho d\phi d\theta.$$

Thus, expressing problem (17) in spherical coordinates yields the following formulation for every $V \in H_0^1(\Sigma \times (0,T))$:

$$\begin{aligned}
\min\ & E(\sigma, a, T) = \frac{1}{2}\int_{\Sigma} (|\dot{u}(\rho,\varphi,\theta,T)|^2 + (\frac{\partial u}{\partial \rho})^2 + \frac{1}{\rho^2}(\frac{\partial u}{\partial \varphi})^2 \\
& + \frac{1}{\rho^2 \sin^2\varphi}(\frac{\partial u}{\partial \theta})^2)\rho^2 \sin\varphi\, d\rho d\phi d\theta \\
subject\ to: & \int_{\Sigma} \dot{u}(T)V(T)\,\rho^2 \sin\varphi\, d\rho d\phi d\theta - \int_{\Sigma} u(T)\dot{V}(T)\,\rho^2 \sin\varphi\, d\rho d\phi d\theta \\
& + \int_0^T \int_{\Sigma} u\ddot{V}\,\rho^2 \sin\varphi\, d\rho d\phi d\theta dt + \int_0^T \int_{\Sigma} \nabla u.\nabla V\,\rho^2 \sin\varphi\, d\rho d\phi d\theta dt \\
& + \int_{\Sigma} a(\rho,\varphi,\theta)u(T)V(T)\,\rho^2 \sin\varphi\, d\rho d\phi d\theta \\
& - \int_{\Sigma} a(\rho,\varphi,\theta)u_0(\rho,\varphi,\theta)V(0)\,\rho^2 \sin\varphi\, d\rho d\phi d\theta \\
& - \int_0^T \int_{\Sigma} a(\rho,\varphi,\theta)u\dot{V}\,\rho^2 \sin\varphi\, d\rho d\phi d\theta dt = Y, \\
& \int_{\sigma} \rho^2 \sin\varphi\, d\rho d\phi d\theta = L\int_{\Sigma} \rho^2 \sin\varphi\, d\rho d\phi d\theta, \quad 0 < L < 1,
\end{aligned} \tag{18}$$

where

$$Y = \int_{\Sigma} (u_1(\rho,\varphi,\theta)V(0) - u_0(\rho,\varphi,\theta)\dot{V}(0))\,\rho^2 \sin\varphi d\rho d\phi d\theta.$$

We integrate the cost and constraint functions in (18) over $[0,T]$. Therefore, this problem can be represented in a new variational form for every $V \in H_0^1(\Sigma \times (0,T))$ as follows:

$$\begin{aligned}
\min\ & E(\sigma, a, T) = \frac{1}{2T}\int_0^T \int_{\Sigma} (|\dot{u}(\rho,\varphi,\theta,T)|^2 + (\frac{\partial u}{\partial \rho})^2 + \frac{1}{\rho^2}(\frac{\partial u}{\partial \varphi})^2 \\
& + \frac{1}{\rho^2 \sin^2\varphi}(\frac{\partial u}{\partial \theta})^2)\rho^2 \sin\varphi\, d\rho d\phi d\theta dt \\
subject\ to: & \frac{1}{T}(\int_0^T \int_{\Sigma} \dot{u}(T)V(T)\,\rho^2 \sin\varphi\, d\rho d\phi d\theta dt
\end{aligned} \tag{19}$$

$$
\begin{aligned}
&- \int_0^T \int_\Sigma u(T)\dot{V}(T)\,\rho^2 \sin\varphi \, d\rho d\phi d\theta dt) \\
&+ \int_0^T \int_\Sigma u\ddot{V}\,\rho^2 \sin\varphi \, d\rho d\phi d\theta dt + \int_0^T \int_\Sigma \nabla u.\nabla V\,\rho^2 \sin\varphi \, d\rho d\phi d\theta dt \\
&+ \frac{1}{T}\int_0^T \int_\Sigma a(\rho,\varphi,\theta)u(T)V(T)\,\rho^2 \sin\varphi \, d\rho d\phi d\theta dt \\
&- \frac{1}{T}\int_0^T \int_\Sigma a(\rho,\varphi,\theta)u_0(\rho,\varphi,\theta)V(0)\,\rho^2 \sin\varphi \, d\rho d\phi d\theta dt \\
&- \int_0^T \int_\Sigma a(\rho,\varphi,\theta)u\dot{V}\,\rho^2 \sin\varphi \, d\rho d\phi d\theta dt = Y, \\
&\int_\sigma \rho^2 \sin\varphi \, d\rho d\phi d\theta - \frac{L}{T}\int_0^T \int_\Sigma \rho^2 \sin\varphi \, d\rho d\phi d\theta dt = 0.
\end{aligned} \tag{20}
$$

In fact, regarding the new formulation in (19), a 3D shape σ is designed as the damping set in spherical coordinates which is smooth, bounded with a specified volume and placed over the $\varphi = \frac{\pi}{2}$ plane. It is also supposed that the boundary of σ consists of two parts. The first one is an unknown smooth surface S, which is defined by the equation $\varphi = \zeta(\rho,\theta) \in C^2(\mathbb{R}^2)$ and the second part is an unknown region D, which is the projection of S in the $\varphi = \frac{\pi}{2}$ plane (see Figure 1). In addition, a simple smooth closed curve ∂D, which is represented by the piecewise smooth function $r = h(\theta),\ 0 \le \theta \le 2\pi$ in polar coordinates and passes through a known point $B = (h(\theta_0),\ \frac{\pi}{2},\ \theta_0)$ surrounds D; see Figure 2.

We recall that there exist the following relations for D and ∂D in this problem:

$$
D = \left\{ (\rho, \frac{\pi}{2}, \theta) \,|0 \le \theta \le 2\pi, \quad 0 \le \rho \le h(\theta) \right\}, \tag{21}
$$

$$
\partial D = \left\{ (\rho, \frac{\pi}{2}, \theta) \,|0 \le \theta \le 2\pi, \quad \rho = h(\theta) \right\}.
$$

Here, an unknown smooth surface S which is defined by the equation $\varphi = \zeta(\rho,\theta) \in C^2(\mathbb{R}^2)$ is called admissible and S_A is considered the set containing all such admissible surfaces. Also, since shape σ is bounded, we conclude that there exist scalars $\varphi_{\min}$, $\varphi_{\max}$ and $\rho_{\max}$ such that, $\varphi_{\min} \le \varphi \le \varphi_{\max}$ and $0 \le \rho \le \rho_{\max}$.

5 Transferring the OSD problem to a classical optimal control problem

In solving problem (19) by applying a new version of the shape-measure method, first, it must be transformed into a classical optimal control problem. For this purpose, the functions (φ, u)

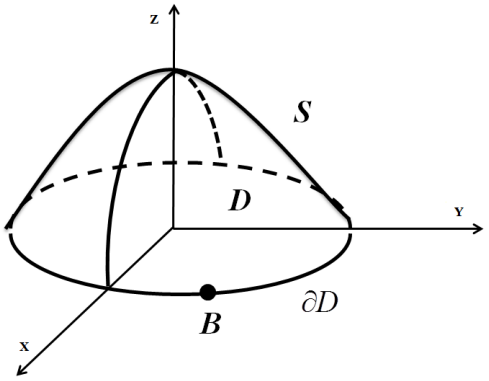


Figure 1: A version of unknown shape.

are considered the trajectory (state) and the vectors $(\dot{u}, u_\rho, u_\varphi, u_\theta, \zeta_\rho, \zeta_\theta, \zeta_{\rho\theta})$ are considered the control functions (the necessity of defining artificial controls ζ_ρ, ζ_θ and $\zeta_{\rho\theta}$ will be explained in later sections). The first constraint in (19) specifies the relation between variables u, $\dot{u}$, u_ρ, u_φ and u_θ. In addition, the dependency of the artificial controls ζ_ρ, ζ_θ and $\zeta_{\rho\theta}$ is demonstrated as follows: Assume that $h_1(\varphi, \theta)$ and $h_2(\rho, \varphi)$ are two arbitrary functions of class $C^1(D \times A)$ (the space of continuously differentiable functions on $D \times A$), by using the chain rule (we remember that $\varphi = \zeta(\rho, \theta)$), we have

$$\frac{\partial h_1}{\partial \rho} = \frac{\partial h_1}{\partial \varphi}\frac{\partial \varphi}{\partial \rho} = \frac{\partial h_1}{\partial \varphi}\zeta_\rho, \quad \frac{\partial h_2}{\partial \theta} = \frac{\partial h_2}{\partial \varphi}\frac{\partial \varphi}{\partial \theta} = \frac{\partial h_2}{\partial \varphi}\zeta_\theta. \tag{22}$$

Definition 1. We define the vector $P = (\varphi, u, \dot{u}, u_\rho, u_\varphi, u_\theta, \zeta_\rho, \zeta_\theta, \zeta_{\rho\theta})$ as admissible if the following conditions hold:

(i) φ is an absolutely continuous trajectory function that takes its values in a bounded subset of $\mathbb{R}$.

(ii) u is the weak solution of system (1) that corresponds to (15) and takes its values in a bounded set U.

(iii) The control functions $\dot{u}$, u_ρ, u_φ, u_θ, ζ_ρ, ζ_θ and $\zeta_{\rho\theta}$ are piecewise-continuous and take their values in the bounded subsets $\dot{U}$, U_ρ, U_φ, U_θ, F_ρ, F_θ and $F_{\rho\theta}$ in $\mathbb{R}$. Moreover, the control functions ζ_ρ, ζ_θ and $\zeta_{\rho\theta}$ should be Lebesgue measurable.

(iv) Vector P satisfies constraints (19) and (22).

By considering the regularity of the initial data (u_0, u_1), as explained in Section 2, the existence of a weak solution for (1) is guaranteed. Consequently, the set of all the admissible

vectors is nonempty, which is denoted by W. By Definition 1, each admissible surface $S \in S_A$ can be replaced exactly by one admissible vector $P = (\varphi, u, \dot{u}, u_\rho, u_\varphi, u_\theta, \zeta_\rho, \zeta_\theta, \zeta_{\rho\theta}) \in W$ [14]. In this regard, minimizing the objective function of (19) over S_A is equivalent to minimizing it over W. Therefore, we can convert problem (19) to a classical optimal control problem.
To embed the solution space, we define regions D' and D'' as follows:

$$\begin{aligned} D^{'} &= [0,T] \times [0,2\pi] \times [0,\rho_\Sigma] \times [0,\varphi_\Sigma] \times U \times \dot{U} \times U_\rho \times U_\varphi \times U_\theta, \\ D^{''} &= D_1 \times F_\rho \times F_\theta \times F_{\rho\theta}, \end{aligned} \tag{23}$$

where $D_1 = \left(D \times \left\{\frac{\pi}{2}\right\}\right) \cup \left([0,\ \rho_{\max}] \times [0,\ 2\pi\ ] \times \left[0,\ \frac{\pi}{2}\right]\right)$ and F_ρ, F_θ and $F_{\rho\theta}$ are bounded subsets of $\mathbb{R}$ which are defined in Definition 1.

6 Embedding the solution space

With the help of the shape-measure approach, which is based on an embedding process, first, we reduce the problem to one consisting of minimizing a linear function over a set of positive measures. Afterwards, we approximate the resulting problem using a finite-dimensional linear programming problem for finding the nearly optimal solution. Then, we use the solution of this problem to construct a sufficiently accurate approximate solution for the original problem. As mentioned, the remarkable advantages of this method over other methods include linear treatment for nonlinear shape optimization problems and global optimization independent of the initial shape and mesh design. Some of the most important existing issues regarding this optimal control problem are as follows:

(i) The possibility of lacking a minimizing admissible vector for E in W.

(ii) Characterizing an admissible minimizer vector for E in W could be very difficult, and even it might be more difficult or impossible to estimate numerically.

It appears that these situations may become more favorable if set W could somehow be made larger. Therefore, we apply the transformation for embedding (19) into another nonclassical problem that has better properties in some respects, allowing us to overcome the mentioned difficulties.

6.1 Characteristics of admissible vectors

To solve problem (19) using the embedding process [28], first, we point out some characteristics of the admissible vector in W. Assume that B is an open ball containing $D \times A$ (region D is defined in (21)) and that $C'(B)$ is the space of real-valued continuously differentiable functions such that their first and second derivatives are continuous and bounded on B, for $\Psi \in C'(B)$, we define the function $\Psi^g \in C(D'')$ as

$$\Psi^g(\rho,\varphi,\theta,\zeta_\rho,\zeta_\theta,\zeta_{\rho\theta}) = (\Psi_{\theta\varphi} - \Psi_\theta \cot\varphi)\frac{\zeta_\rho}{\zeta_\theta} - \Psi_{\rho\varphi} - \frac{1}{\rho}\Psi_\varphi.$$

The function Ψ has continuous partial derivatives and according to Definition 1, the control functions ζ_ρ, ζ_θ and $\zeta_{\rho\theta}$ are Lebesgue measurable. As a result, Ψ^g is integrable on $D \times A$ [29] and by using Stokes' theorem [32], for $F = \nabla(\rho\,\Psi)$, we have

$$\begin{aligned}\oint_{\partial S} F dR &= \iint_S \nabla \times F.\, n\, d\sigma = \iint_D \nabla \times F.\, \frac{\nabla K}{|\nabla K|}\frac{|\nabla K|}{|\nabla K.p|} dA \\ &= \iint_D ((\Psi_{\theta\varphi} - \Psi_\theta \cot\varphi)\frac{\zeta_\rho}{\zeta_\theta} - \Psi_{\rho\varphi} - \frac{1}{\rho}\Psi_\varphi)\rho d\rho d\theta \equiv \rho_0 \Delta_\Psi, \end{aligned} \tag{24}$$

where $\rho_0 \equiv h(\theta_0)$ and since ∂D is a closed curve, $\Delta_\Psi = \Psi(\rho_0, \frac{\pi}{2}, 2\pi) - \Psi(\rho_0, \frac{\pi}{2}, 0)$. Additionally, Ψ_θ, Ψ_φ and $\Psi_{\theta\varphi}$, $\Psi_{\rho\varphi}$ are the first and second order partial derivatives of Ψ.

The curl of the arbitrary function $F = F_\rho\hat{\rho} + F_\varphi\hat{\varphi} + F_\theta\hat{\theta}$ in spherical coordinates is defined from [6] as follows:

$$\begin{aligned} curl\, F &= \frac{1}{\rho^2 \sin\varphi}\begin{vmatrix} \hat{\rho} & \rho\hat{\varphi} & \rho\sin(\varphi)\hat{\theta} \\ \frac{\partial}{\partial\rho} & \frac{\partial}{\partial\varphi} & \frac{\partial}{\partial\theta} \\ F_\rho & \rho F_\varphi & \rho\sin(\varphi)F_\theta \end{vmatrix} \\ &= \frac{1}{\rho\sin\varphi}(\frac{\partial}{\partial\varphi}(F_\theta\sin\varphi) - \frac{\partial F_\varphi}{\partial\theta})\hat{\rho} \\ &\quad + \frac{1}{\rho}(\frac{1}{\sin\varphi}\frac{\partial F_\rho}{\partial\theta} - \frac{\partial}{\partial\rho}(\rho F_\theta))\hat{\varphi} + \frac{1}{\rho}(\frac{\partial}{\partial\rho}(\rho F_\varphi) - \frac{\partial F_\rho}{\partial\varphi})\hat{\theta}, \end{aligned}$$

where $\hat{\rho}, \hat{\varphi}$ and $\hat{\theta}$ are the unit vectors in increasing directions of ρ, φ and θ, respectively and F_ρ, F_φ and F_θ are the components of field F.

Assume that $\wp(D^\circ)$ is the space of infinitely differentiable real-valued functions with compact support in D° (the set of interior points of D). We consider $\gamma(\rho,\theta) \in \wp(D^\circ)$ and define $\Theta \in C(D'')$ as follows:

$$\begin{aligned}\Theta(\rho,\varphi,\theta,\zeta_\rho,\zeta_\theta,\zeta_{\rho\theta}) =&(2+\frac{(\varphi-\frac{\pi}{2})\cos\varphi}{\sin\varphi}-\frac{1}{\sin\varphi}-\frac{1}{\rho\sin\varphi})\frac{\zeta_\rho\gamma_\theta\sin\varphi}{\zeta_\theta}\\&+(\frac{-2}{\rho}+\sin\varphi+1)\gamma_\rho\\&+\frac{(\frac{-1}{\rho}+\sin\varphi)}{\zeta_\theta}(\zeta_{\rho\theta}\gamma+\gamma_{\rho\theta}(\varphi-\frac{\pi}{2}))\\&+(\frac{\sin\varphi+1}{\rho}+\zeta_\rho\cos\varphi)\gamma+\frac{1}{\rho\zeta_\theta}\sin(\varphi)(\varphi-\frac{\pi}{2})\gamma_\theta.\end{aligned}$$

Here we have γ_θ and γ_ρ as the first-order partial derivatives of γ and $\gamma_{\rho\theta}$ as the second-order partial derivative of γ. Since every infinitely differentiable function is absolutely continuous because it satisfies the Lipschitz condition, by Definition 1, the trajectory function φ is absolutely continuous, and $(\varphi-\frac{\pi}{2}),\gamma(\rho,\theta)$ is also an absolutely continuous function. Therefore, its partial derivatives are analytically integrable [29].

Now, if we consider $\Psi_1(\rho,\varphi,\theta)=(\varphi-\frac{\pi}{2})\gamma(\rho,\theta)$ and define

$$F\equiv(\Psi_{1\rho},\Psi_{1\phi},\Psi_{1\theta})=(\zeta_\rho\gamma+\gamma_\rho(\varphi-\frac{\pi}{2}),\gamma,\zeta_\theta\gamma+(\varphi-\frac{\pi}{2})\gamma_\theta),$$

then by Stokes' theorem we have

$$\begin{aligned}\oint_{\partial S}F.dR&=\iint_S\nabla\times F.n\,d\sigma=\iint_D\frac{(\nabla\times F)\cdot\nabla K}{|\nabla K\cdot p|}\,dA\\&=\iint_D\frac{(\nabla\times F)\cdot\nabla K}{\frac{\zeta_\theta}{\rho sin\varphi}}\,dA=\iint_D\Theta\,\rho\,d\rho d\theta=0.\end{aligned}\tag{25}$$

The value of γ on ∂D is zero because function γ has compact support in D° [29], and regarding the definition of F, the value of F on ∂D is also equal to zero. Additionally, assuming that $C_1(D'')$ is the space of all functions in $C(D'')$ depending only on the variables ρ and θ, we have

$$\int_S\frac{\zeta_\theta}{\sqrt{\zeta_\rho^2+\frac{1}{\rho^2}+\frac{\zeta_\theta^2}{\rho^2\sin^2\varphi}}\,\rho\sin\varphi}g(\rho,\theta)\,d\sigma=\iint_D g(\rho,\theta)\rho d\rho d\theta\equiv a_g,\tag{26}$$

for all $g\in C_1(D'')$, where a_g is the Lebesgue integral of $g(\rho,\theta)$ over D.

6.2 Transformation

Suppose that D' and D'' are defined as in (23). For any admissible vector P in W, we define the linear, positive functionals Γ_p and Υ_p on $C(D')$ and $C(D'')$ of all continuous real-valued

functions as

$$\Gamma_P(F) = \int_0^T \int_\Sigma F\ \rho^2 \sin\varphi d\rho d\phi d\theta dt, \quad \text{for all } F \in C(D'),$$
$$\Upsilon_P(I) = \iint_D I\rho d\rho d\theta, \qquad \text{for all } I \in C(D'').$$

To enlarge the set W, we denote the space of all positive Radon measures on D' and D'' by $M^+(D')$ and $M^+(D'')$, respectively. Using Riesz's representation theorem [29], there exist unique positive Radon measures $\mu \in M^+(D')$ and $\upsilon \in M^+(D'')$ such that

$$\Gamma(F) = \int_0^T \int_\Sigma F\, d\mu \equiv \mu(F), \quad \text{for all } F \in C(D'), \tag{27}$$

$$\Upsilon(I) = \iint_D I\, d\lambda \equiv \upsilon(I), \quad \text{for all } I \in C(D''). \tag{28}$$

According to [28], these transformations are injective. Consequently, any admissible element can be displayed as (27) or (28) by a unique pair of measures, say (μ, υ), in a subset W of $M^+(D') \times M^+(D'')$. Therefore, one can transform problem (19) into a measure space as follows:

$$(\varphi, u, \dot{u}, u_\rho, u_\varphi, u_\theta, \zeta_\rho, \zeta_\theta, \zeta_{\rho\theta}) \in P \longmapsto (\mu, \upsilon) \in M^+(D') \times M^+(D'').$$

Given the definitions of the measures in (27) and (28), it is necessary to rewrite constraint (16) with respect to only one of the corresponding integrals, μ or υ. To this end, and following [6, 32], we can apply the divergence theorem (see Theorem 1) by regarding

$$\nabla K = (\zeta_\rho, \frac{-1}{\rho}, \frac{1}{\rho\sin\varphi}\zeta_\theta),$$
$$dS = \frac{|\nabla K|}{|\nabla K.p|} dA = \frac{\sqrt{\zeta_\rho^2 + \frac{1}{\rho^2} + \frac{1}{\rho^2\sin^2\varphi}\zeta_\theta^2}}{\zeta_\theta} \rho\sin(\varphi)\rho d\rho d\theta.$$

Suppose that the equation of surface S is given by $K \equiv \zeta(\rho, \theta) - \varphi = 0$, where dS is the differential element of the surface area and $p = (0, 0, 1)$ is a unit vector normal to D. It is important to mention that the area element is given by $dA = \rho\sin(\varphi)d\rho d\theta$. Thus, we have $dA = \rho d\rho d\theta$ because in D we have $\varphi = \frac{\pi}{2}$. Hence, the calculation can be simplified by converting the integral over σ in (19) into an integral over D with $\partial\sigma = S \bigcup D$ using the divergence theorem [32] as follows:

$$\int_\sigma div\, F dV = \oint_{S \bigcup D} F.n\, dS$$

$$= \iint\limits_D F.\frac{\nabla K}{|\nabla K|}\frac{\sqrt{\zeta_\rho^2+\frac{1}{\rho^2}+\frac{1}{\rho^2\sin^2\varphi}\zeta_\theta^2}}{\zeta_\theta}\rho\sin(\varphi)\rho d\rho d\theta,$$

where n is an outward-pointing unit normal vector of the boundary σ. Since the gradient vector of K in spherical coordinates is $\nabla K = (\zeta_\rho, \frac{-1}{\rho}, \frac{1}{\rho\sin\varphi}\zeta_\theta)$, we have $|\nabla K| = \sqrt{\zeta_\rho^2+\frac{1}{\rho^2}+\frac{1}{\rho^2\sin^2\varphi}\zeta_\theta^2}$. Also, $div\ F$ in spherical coordinates is calculated by the following formula:

$$\frac{1}{\rho^2}\frac{\partial}{\partial\rho}\left(\rho^2F_\rho\right)+\frac{1}{\rho\sin\varphi}\left(\frac{\partial F_\theta}{\partial\theta}\right)+\frac{1}{\rho\sin\varphi}\frac{\partial}{\partial\varphi}\left(\sin\varphi F_\varphi\right).$$

Because $\int\limits_\sigma dV$ is the volume constraint, we consider the vector function F so that its divergence is a constant number. If we substitute $F = \rho\hat{\rho}+\rho\hat{\varphi}-\theta\rho\cos(\varphi)\hat{\theta}$ in the above statement, from the definition of divergence in spherical coordinates [6], we have $div\,F = 3$ (with $\hat{\rho}$, $\hat{\varphi}$ and $\hat{\theta}$ as unit vectors in increasing directions of ρ, φ and θ, respectively). We conclude

$$\begin{aligned}
&\iint\limits_D F.\frac{\nabla K}{|\nabla K|}\frac{\sqrt{\zeta_\rho^2+\frac{1}{\rho^2}+\frac{1}{\rho^2\sin^2\varphi}\zeta_\theta^2}}{\zeta_\theta}\rho\sin(\varphi)\rho d\rho d\theta\\
&= \iint\limits_D (\rho,\rho,-\theta\rho\cos(\varphi)).\left(\frac{\left(\zeta_\rho,\frac{-1}{\rho},\frac{1}{\rho\sin\varphi}\zeta_\theta\right)}{\sqrt{\zeta_\rho^2+\frac{1}{\rho^2}+\frac{1}{\rho^2}\sin^2\varphi\zeta_\theta^2}}\right)\\
&\quad\frac{\sqrt{\zeta_\rho^2+\frac{1}{\rho^2}+\frac{1}{\rho^2}\sin^2\varphi\zeta_\theta^2}}{\zeta_\theta}\rho\sin(\varphi)\rho d\rho d\theta.
\end{aligned}$$

As a result,

$$\int\limits_\sigma dV = \frac{1}{3}\iint\limits_D\frac{(\rho\zeta_\rho-1-\theta\zeta_\theta\cot\varphi)\rho\sin\varphi}{\zeta_\theta}\rho d\rho d\theta. \tag{29}$$

Therefore, the volume constraint (the last constraint in (19)) would be rewritten in the following way:

$$\frac{1}{3}\iint\limits_D\frac{(\rho\zeta_\rho-1-\theta\zeta_\theta\cot\varphi)\rho\sin\varphi}{\zeta_\theta}\rho d\rho d\theta = \frac{L}{T}\int_0^T\int_\Sigma\rho^2\sin\varphi\,d\rho d\phi d\theta dt.$$

6.3 Converting the problem into a measure space

Regarding relations (22), (24), (25), (26) and (29), the known properties of admissible elements of P and the definitions of measures μ and υ in (27) and (28), problem (19) can now be expressed as follows based on unknown measures υ and μ:

$$\begin{aligned}
\min\ E(\mu,\upsilon) &= \frac{1}{2T}\mu(|\dot{u}(\rho,\varphi,\theta,T)|^2 + (\frac{\partial u}{\partial \rho})^2 \\
&\quad + \frac{1}{\rho^2}(\frac{\partial u}{\partial \varphi})^2 + \frac{1}{\rho^2 \sin^2\varphi}(\frac{\partial u}{\partial \theta})^2) \\
\textit{subject to}: \quad & \upsilon(\Psi^g) = \rho_0 \Delta_\Psi, \quad \text{for all } \Psi \in C'(B), \\
& \upsilon(\Theta) = 0, \quad \text{for all } \gamma \in \wp(D^0), \\
& \upsilon(g(\rho,\theta)) = a_g, \quad \text{for all } g \in C_1(D''), \\
& \frac{1}{T}\mu(\dot{u}(T)V(T) - u(T)\dot{V}(T)) + \mu(u\ddot{V} - au\dot{V}) \\
& - \mu(\nabla u.\nabla V) + \frac{1}{T}\mu(au(T)V(T)) - \frac{1}{T}\mu(au_0V(0)) \\
& \quad = Y, \quad \text{for all } V \in H_0^1(\Sigma \times (0,T)), \\
& \upsilon(F_0) - \frac{3L}{T}\mu(1) = 0, \\
& \upsilon(L_1) = 0, \quad \text{for all } h_1(\varphi,\theta) \in C^1(D \times A), \\
& \upsilon(L_2) = 0, \quad \text{for all } h_2(\rho,\varphi) \in C^1(D \times A),
\end{aligned} \tag{30}$$

where $L_1 = \frac{\partial h_1}{\partial \rho} - \frac{\partial h_1}{\partial \varphi}\zeta_\rho$, $L_2 = \frac{\partial h_2}{\partial \theta} - \frac{\partial h_2}{\partial \varphi}\zeta_\theta$, $F_0 = \frac{(\rho\zeta_\rho - 1 - \theta\zeta_\theta \cot\varphi)\rho\sin\varphi}{\zeta_\theta}$, and

$$Y = \int_\Sigma (u_1(\rho,\varphi,\theta)V(0) - u_0(\rho,\varphi,\theta)\dot{V}(0))\,\rho^2 \sin\varphi d\rho d\phi d\theta.$$

We recall that the theoretical measure problem (30) is linear with respect to the two measures μ and υ even though the initial problem is nonlinear.

The space $M^+(D') \times M^+(D'')$ is a linear space which will become a locally convex topological vector space when the weak* topology is given [28].

The proof of the following existence theorem can be found in [8, 3, 28].

Theorem 4. a) $\tilde{Q} \subseteq M^+(D') \times M^+(D'')$ is compact under the weak* topology on $M^+(D') \times M^+(D'')$

b) The objective function $E(\mu,\upsilon)$ of problem (30) is continuous.

c) There exists a pair of measures (μ^*, υ^*) which is an optimal solution of problem (30) in the set $\tilde{Q} \subset M^+(D') \times M^+(D'')$; that is, for every $(\mu, \upsilon) \in \tilde{Q}$, we have

$$E(\mu^*, \upsilon^*) \leq E(\mu, \upsilon).$$

Even though (30) has an optimal solution in $\tilde{Q}$, it is still very difficult to obtain the exact solution due to the fact that the underlying spaces are not finite-dimensional because the number of equations is not finite and the unknowns are measures. Therefore, it is completely acceptable to search for a suboptimal solution. Thus, first, by choosing suitable dense subsets in the appropriate spaces and then, by choosing a finite number of them, the problem is approximated by semi-finite linear programming.

7 Approximation schemes

Measure-theoretical problem (30) is an infinite-dimensional LPP and hence, it is not easy to solve. In this section, we shall apply approximation schemes to find a nearly optimal solution of (30) by solving a finite-dimensional linear program. In this regard, first the function $(\mu, \upsilon) \to E(\mu, \upsilon)$ should be minimized, but over a subset of $\tilde{Q}$ that satisfies only a finite number of constraints of (30) called $\tilde{Q}(M_1, M_2, R_1, R_2, M_3, M_4, M_5)$. To achieve this goal, we select countable sets of functions whose linear combinations are dense in the appropriate spaces. Therefore, we consider total function sets $\{\Psi_i, i \in N\}$, $\{\gamma_h, h \in N\}$, $\{g_{sk}, s, k \in N\}$, $\{V_j, j \in N\}$ and $\{h_{1l_1}, h_{2l_2}, l_1, l_2 \in N\}$ in spaces $C'(B)$, $\wp(D^\circ)$, $C_1(D'')$, $H_0^1(\Sigma \times (0, T))$ and $C^1(D \times A)$, respectively and select a finite number of them to achieve the following approximated problem for (30):

$$\begin{aligned}
\min E(\mu, \upsilon) = {} & \frac{1}{2T}\mu(|\dot{u}(\rho, \varphi, \theta, T)|^2 + (\frac{\partial u}{\partial \rho})^2 \\
& + \frac{1}{\rho^2}(\frac{\partial u}{\partial \varphi})^2 + \frac{1}{\rho^2 \sin^2 \varphi}(\frac{\partial u}{\partial \theta})^2) \\
\textit{subject to}: \quad & \upsilon(\Psi_i^g) = \rho_0 \Delta_{\Psi i}, \quad i = 1, 2, \ldots, M_1, \\
& \upsilon(\Theta_h) = 0, \qquad h = 1, 2, \ldots, M_2, \\
& \upsilon(g_{sk}(\rho, \theta)) = a_{g_{sk}}, \ s = 1, 2, \ldots, R_1, \ k = 1, 2, \ldots, R_2, \\
& \frac{1}{T}\mu(\dot{u}(T)V_j(T) - u(T)\dot{V}_j(T)) + \mu(u\ddot{V}_j - au\dot{V}_j) \\
& - \mu(\nabla u.\nabla V_j) + \frac{1}{T}\mu(au(T)V_j(T)) - \frac{1}{T}\mu(au_0 V_j(0))
\end{aligned}$$

$$
\begin{aligned}
&= Y,\ j = 1, 2, \ldots, M_3,\\
&\upsilon(F_0) - \frac{3L}{T}\mu(1) = 0,\\
&\upsilon(L_{1l_1}) = 0, \qquad l_1 = 1, 2, \ldots, M_4,\\
&\upsilon(L_{2l_2}) = 0, \qquad l_2 = 1, 2, \ldots, M_5.
\end{aligned} \tag{31}
$$

According to [28], if M_1, M_2, R_1, R_2, M_3, M_4 and M_5 tend to infinity, the solution of problem (31) is a very good approximation for the main problem (30). Also, from [28], we can characterize the pair of optimal measures of (27)–(28) as

$$
\mu^* = \sum_{n=1}^{M_3+1} \hat{\alpha}_n^* \delta(Z_n^*),\ \upsilon^* = \sum_{m=1}^{M_1+M_2+(R_1\times R_2)+M_4+M_5+1} \hat{\beta}_m^* \hat{\delta}(z_m^*), \tag{32}
$$

with $\hat{\alpha}_n^*$, $\hat{\beta}_m^* \geq 0$ and $Z_n^* \in D'$, $z_m^* \in D''$. Here $\delta(t)$ and $\hat{\delta}(t)$ are unitary atomic measure supported by singleton sets $\{t\}$. Then, if we substitute (32) in (31), it becomes equivalent to a nonlinear optimization problem whose unknowns are the coefficients $\hat{\alpha}_n^*$, $\hat{\beta}_m^*$ and their supports $\{Z_n^*\}$ and $\{z_m^*\}$, $n = 1, 2, \ldots, M_3 + 1$ and $m = 1, 2, \ldots, M_1 + M_2 + (R_1 \times R_2) + M_4 + M_5 + 1$.

If we approximate the supports $\{Z_n^*\}$ and $\{z_m^*\}$ by introducing the support points in a countable dense subset of D' and D'', respectively (referred to as W_1 and W_2), we will obtain the corresponding finite LP as desired. Therefore, one can transform (31) into a finite linear programming problem with unknown coefficients $\hat{\alpha}_n$ and $\hat{\beta}_m$ as follows:

$$
\begin{aligned}
\min\ E(\hat{\alpha},\, \hat{\beta},\, D) = {} & \frac{1}{2T}\sum_{n=1}^{N_1} \hat{\alpha}_n (|\dot{u}(Z_n)|^2 + \Big(\frac{\partial u(Z_n)}{\partial \rho_n}\Big)^2\\
& + \frac{1}{{\rho_n}^2}\Big(\frac{\partial u(Z_n)}{\partial \varphi_n}\Big)^2 + \frac{1}{{\rho_n}^2 \sin^2\varphi_n}\Big(\frac{\partial u(Z_n)}{\partial \theta_n}\Big)^2)\\
subject\,to: \quad & \sum_{m=1}^{N_2} \hat{\beta}_m \Psi_i^g(z_m) = \rho_0 \Delta_{\Psi i}, \qquad i = 1, 2, \ldots, M_1,\\
& \sum_{m=1}^{N_2} \hat{\beta}_m \Theta_h(z_m) = 0, \qquad h = 1, 2, \ldots, M_2,\\
& \sum_{m=1}^{N_2} \hat{\beta}_m g_{sk}(z_m) = 0,\ s = 1, 2, \ldots, R_1,\ k = 1, 2, \ldots, R_2,\\
& \frac{1}{T}\sum_{n=1}^{N_1} \hat{\alpha}_n (\dot{u}(Z_n) V_j(Z_n) - u(Z_n)\dot{V}_j(Z_n))\\
& + \sum_{n=1}^{N_1} \hat{\alpha}_n (u(Z_n)\ddot{V}_j(Z_n) - a u(Z_n)\dot{V}_j(Z_n))
\end{aligned}
$$

$$
\begin{aligned}
&-\sum_{n=1}^{N_1}\hat{\alpha}_n\nabla u(Z_n).\nabla V_j(Z_n)+\frac{1}{T}\sum_{n=1}^{N_1}\hat{\alpha}_n a u(Z_n)V_j(Z_n)\\
&-\frac{1}{T}\sum_{n=1}^{N_1}\hat{\alpha}_n a u_0(Z_n)V_j(0)=Y,\ j=1,\,2,\,\ldots,\,M_3,\\
&\sum_{m=1}^{N_2}\hat{\beta}_m F_0(z_m)-\frac{3L}{T}\sum_{n=1}^{N_1}\hat{\alpha}_n=0,\\
&\sum_{m=1}^{N_2}\hat{\beta}_m L_{1l_1}(z_m)=0,\qquad l_1=1,\,2,\,\ldots,\,M_4,\\
&\sum_{m=1}^{N_2}\hat{\beta}_m L_{2l_2}(z_m)=0,\qquad l_2=1,\,2,\,\ldots,\,M_5,\\
&\hat{\alpha}_n\geq 0,\ n=1,2,\ldots,M_3+1,\\
&\hat{\beta}_m\geq 0,\ m=1,2,\ldots,M_1+M_2+(R_1\times R_2)+M_4+M_5+1,
\end{aligned}
\tag{33}
$$

where W_1 and W_2 are the countable dense subsets of D' and D'', respectively, and N_1 and N_2 are sufficiently large numbers. Sets $\{Z_n,\ n=1,2,\ldots,M_3+1\}\subset W_1$ and $\{z_m,\ m=1,m=2,\ldots,M_1+M_2+(R_1\times R_2)+M_4+M_5+1\}\subset W_2$ are constructed by dividing appropriate intervals in the definitions of D' and D'' (see (23)) into a finite number of equal subintervals, respectively.

On the other hand, we know that a curve can usually be approximated with broken lines. So suppose that ∂D is approximated by a number M of its points (corners of broken lines belonging to ∂D) which will be called the M-representation of D. We can assign fixed -components such as $\theta_i=\theta'_i$, $i=1,\,2,\,\ldots,\,M$, to the points in the M-representation set to obtain a fixed number M for shape variables, without loss of generality. Therefore, we can characterize each admissible M-representation set, denoted as D_M, by $\rho_1\sin\varphi_1,\ \rho_2\sin\varphi_2,\ \ldots,\ \rho_M\sin\varphi_M$. So, we can determine each region of $D\in D_M$, by a set of finite points $(\theta_m,\ \rho_m\sin\varphi_m)$, $m=1,\,2,\,\ldots,\,M$. Then, we can find the nearly optimal value for $E(\hat{\alpha}^*,\ \hat{\beta}^*,D)$ as a function of variables $\rho_1,\ \rho_2,\ \ldots,\ \rho_M$ and $\varphi_1,\ \varphi_2,\ \ldots,\ \varphi_M$, for a given $D\in D_M$, by solving (33). As a result, we can determine the following vector function:

$$
J:(\rho_1,\ \rho_2,\ \ldots,\ \rho_M,\ \varphi_1,\ \varphi_2,\ \ldots,\ \varphi_M)\in\mathbb{R}^{2M}\mapsto E(\hat{\alpha}^*,\ \hat{\beta}^*,\ D)\in\mathbb{R}. \tag{34}
$$

The global minimizer of this function can be identified using a suitable optimization algorithm, such as PSO, such that each time the algorithm needs to evaluate J, a version of problem (33) must be solved. In this regard, the optimal coefficients $\hat{\alpha}_j^*$ and $\hat{\beta}_j^*$ are characterized. Whenever the function J is minimized, the minimizer $(\rho_1^*,\ \rho_2^*,\ \ldots,\ \rho_M^*,\ \varphi_1^*,\ \varphi_2^*,\ \ldots,\ \varphi_M^*)$ and consequently,

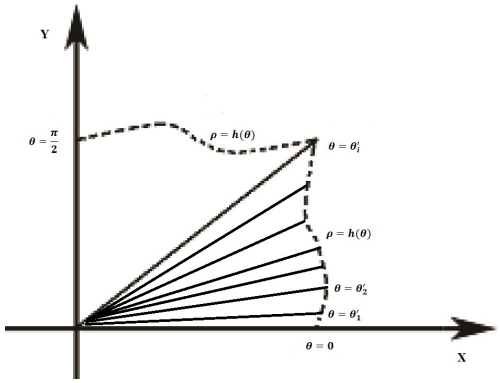


Figure 2: An example of unknown domain D.

the optimal region D^* are obtained. Then, related to the optimal region D^*, optimal points $(\theta_m^*, \rho_m^*, \varphi_m^*)$ corresponding to optimal coefficients $\hat{\beta}_m^* > 0$ are obtained.

We might find some outliers among these optimal points [16] as a result of using approximation scheme (we recall that an outlier is an observation that deviates significantly from other observations). These outlier points may cause some difficulties in achieving the smoothness of the optimal shape. Therefore, they can be rejected from the set of optimal points if we employ a suitable outlier detection algorithm [1], such as the LoOP algorithm in [19]. Then, by applying a smooth curve fitting, we can achieve a smooth surface with a shape that is nearly optimal for the specified damping set.

7.1 Introducing total sets

In this subsection, we will explain how to select total functions for problem (30). Consider M_1 functions $\Psi_i \in C'(B)$ as monomials of the form $\rho^{s_1}\varphi^{s_2}\theta^{s_3}$, where $s_1, s_2, s_3 \in \{0, 1, 2, \ldots\}$, and M_2 functions $\gamma_h \in \wp(D^\circ)$ as

$$\begin{aligned}
\gamma_{1i} &= (\rho - h(\theta))(\sin(i\pi\theta)),\ i = 1, 2, 3, \ldots,\\
\gamma_{2i} &= (\rho - h(\theta))(\cos(i\pi\theta)),\ i = 1, 2, 3, \ldots,\\
\gamma_{3i} &= (\rho - h(\theta))(\sin(i\pi\theta))(\cos(i\pi\theta)),\ i = 1, 2, 3, \ldots .
\end{aligned}$$

where the function $h(\theta)$ represents the boundary of D. Now, we will introduce functions which depend only on ρ and θ. Without loss of generality, we assume that the origin is an interior point of D [12]. A finite number of angles $\theta = \theta_i$, $i = 1, 2, \ldots, M$ from the uniform dense subset

of $[0, 2\pi]$ would be considered such that $0 = \theta_1 \leq \theta_2 \leq \cdots \leq \theta_M = 2\pi$. Then, region D can be divided into $M-1$ parts by half-lines $\theta = \theta_i$. Also, the ith part of $D(i = 1, 2, \ldots, M-1)$, called D_i, can be approximated by a sector of the circle with radius $R_i = \frac{\rho_i \sin\varphi_i + \rho_{i+1}\sin\varphi_{i+1}}{2}$, when $\theta_i \leq \theta \leq \theta_{i+1}$. Then $(\rho_1^*, \rho_2^*, \ldots, \rho_M^*)$ and $(\varphi_1^*, \varphi_2^*, \ldots, \varphi_M^*)$ can be obtained optimally corresponding to $(\theta_1, \theta_2, \ldots, \theta_M)$. Obviously, if the number of angles is large enough, the union of D_i's can effectively approximate D. Now, we define function $g_{sk} \in C_1(D'')$ as

$$g_{sk}(\rho, \theta) = \begin{cases} 1, & \theta \in J_{1s} \quad \text{and} \quad \rho \in J_{2k}, \\ 0, & O.W. \end{cases},$$

where $J_{1s} = \left[\frac{(s-1)\theta_i}{R_1}, \frac{(s)\theta_i}{R_1}\right)$, $s = 1, 2, \ldots, R_1$ and $J_{2k} = \left[\frac{(k-1)R_i}{R_2}, \frac{kR_i}{R_2}\right)$ for $k = 1, 2, \ldots, R_2$. So, we have

$$\begin{aligned} \iint_D g_{sk}(\rho, \theta)\rho d\rho d\theta &= \int_{\rho_{k-1}}^{\rho_k} \int_{\theta_{s-1}}^{\theta_s} \rho d\theta d\rho \\ &= \frac{1}{2}(\theta_s - \theta_{s-1})(\rho_k^2 - \rho_{k-1}^2) \equiv a_{g_{sk}}. \end{aligned}$$

Moreover, M_4 functions $h_1(\varphi, \theta) \in C^1(D \times A)$ are taken as the polynomials with respect to φ and θ, and M_5 functions $h_2(\rho, \varphi) \in C^1(D \times A)$ as products of different powers of ρ and φ. From [28], the linear combinations of these functions are identically dense in appropriate spaces.

In this paper, the PSO algorithm is used in Step 3 of the algorithm (Subsection 7.2) and in Section 8 to minimize the objective function J. Specifically, PSO searches for the optimal coordinates of M boundary points in spherical coordinates to draw the 2D region D_M of the damping set. Compared to other heuristic methods like GA, PSO has several clear advantages. First, it is very simple to implement and needs fewer parameters. Second, it has a faster convergence speed because it updates positions directly without complex operations like crossover and mutation. Third, it requires less computer memory. Finally, PSO works very well for continuous problems, which perfectly fits our task of finding continuous geometric boundaries in a 3D space.

7.2 Algorithm of the proposed approach

We present an algorithmic path for the solution method based on the discussions in Sections 6 and 7 to demonstrate how our method could be applied to solve problem (5). In this regard, using the following five steps helps us identify the nearly optimal shape for the damping set.

Step 1: Select n_1, n_2, n_3, n_4, n_5, n_6, n_7, n_8 and n_9 points in the given sets $[0, T]$, $[0, 2\pi]$, $[0, \rho_\Sigma]$, $[0, \varphi_\Sigma]$, U, $\dot{U}$, U_ρ, U_φ and U_θ, respectively. Thus $N_1 = \prod_{i=1}^{9} n_i$ number of 9-dimensional cells is constructed. Then, an arbitrary point Z_j, $j = 1, 2, \ldots, N_1$, is selected in each cell. Similarly, consider m_1, m_2, m_3, m_4, m_5 and m_6 points in the intervals $[0, 2\pi]$, $\left[0, \frac{\pi}{2}\right]$, $[0, \rho_\sigma]$, F_ρ, F_θ and $F_{\rho\theta}$, respectively. So $N_2 = m_1.m_2.m_3.m_4.m_5.m_6$ number of 6-dimensional cells is determined. In each cell, an arbitrary point z_j, $j = 1, 2, \ldots, N_2$ is selected (note that sets $\dot{U}$, U_ρ, U_φ, U_θ, F_ρ, F_θ and $F_{\rho\theta}$ can be chosen by trial and error).

Step 2: We select functions $\Psi_i(z)$, $\gamma_h(z)$, $g_{s\,k}(z)$, $V_j(z)$, $h_{1v}(z)$ and $h_{2l}(z)$ in appropriate spaces as mentioned in Subsection 7.1, for a given initial region D and for fixed numbers M_1, M_2, R_1, R_2, M_3, M_4 and M_5. Now, we can set up the finite LPP (33) with N_1+N_2 variables and $N' \equiv M_1 + M_2 + (R_1 \times R_2) + M_3 + M_4 + M_5 + 1$ constraints.

Step 3: To solve problem (33), we apply a two-phase optimization approach for the given M and the specified number of iterations. First, we establish function J defined by (34). Then, we minimize it using a standard optimization method like those discussed in [11], such as a random search method or PSO algorithm. The algorithm produces M new points in each iteration as $(\theta_m, \rho_m \sin\varphi_m)$, $m = 1, 2, \ldots, M$ which determine an M-representation region D_M as explained in Section 7. A value for J can be obtained for each D_M by running its related LP in (33). Therefore, we determine the optimal value of the energy for system (1) and the optimal coefficients $\hat{\alpha}^*$'s and $\hat{\beta}^*$'s simultaneously by minimizing J based on the optimal values $(\rho_1^*, \rho_2^*, \ldots, \rho_M^*)$ and $(\varphi_1^*, \varphi_2^*, \ldots, \varphi_M^*)$.

Step 4: We characterize the indices n for the optimal values $\hat{\beta}_1^*, \hat{\beta}_2^*, \ldots, \hat{\beta}_{N_2}^*$ resulting from Step 3 so that the coefficient $\hat{\beta}_n^*$ (related to the n-th extreme point) is positive and takes $\rho = \rho_n^*$, $\theta = \theta_n^*$, $\varphi = \varphi_n^*$, $\zeta_\rho = \zeta_{\rho_n}^*$, $\zeta_\theta = \zeta_{\theta_n}^*$ and $\zeta_{\rho\theta} = \zeta_{\rho\theta_n}^*$ (in fact, ρ_n^*, θ_n^*, φ_n^*, $\zeta_{\rho_n}^*$, $\zeta_{\theta_n}^*$ and $\zeta_{\rho\theta_n}^*$ are the related components associated with point z_n^* corresponding to $\hat{\beta}_n^* > 0$). Consequently, we can determine the nearly optimal shape using points $(\rho_i^*, \varphi_i^*, \theta_i^*)$.

Step 5: We apply the outlier detection algorithm to remove outliers from the set of points obtained in Step 4. Subsequently, to achieve a smooth surface resembling the nearly optimal damping set, we perform a smooth curve fitting using the MATLAB software toolbox on the obtained nearly optimal N' points.

7.3 Theoretical convergence and parameter sensitivity

To address the sensitivity of the proposed method to parameters such as mesh size and the number of basis functions (which determine the number of constraints and variables in the linear

programming formulation), we rely on the theoretical foundations established in the measure-theoretic framework. According to the fundamental work [10] on the linear programming approach to 2D shape design, as both the number of constraints and the number of variables in the embedded linear program increase (corresponding respectively to a smaller mesh size and using more basis functions), the solution of the finite-dimensional LP converges globally to the weak solution of the original infinite-dimensional problem. Based on this concept, in our 3D spherical formulation, this theoretical guarantee fact ensures that reducing the mesh size or increasing the number of functions, strictly improves the convergence to the optimal shape of the damping set. However, increasing the number of constraints and variables makes the linear programming problem much harder to solve. Therefore, in our numerical examples, we chose a good balance between accuracy and computation time. The selected mesh size and number of functions provide a good approximation of the optimal shape. At the same time, they ensure a reasonable computation time and global convergence.

8 Numerical examples

To illustrate the application of the algorithm described in the previous section, this section presents two numerical simulations for problem (1). It is worth noting that the examples presented here are extracted from the 2D example provided in [24].

Example 1. Consider the following 3D damped wave equation, which is defined on a spherical domain Σ centered at the origin with a radius of 2 in the time interval $[0,1]$. For all $(t,\rho,\varphi,\theta) \in (0,T)\times\Sigma$, we have

$$\begin{aligned}
&\ddot{u} - \left(\frac{\partial^2 u}{\partial\rho^2} + \frac{2}{\rho}\frac{\partial u}{\partial\rho} + \frac{1}{\rho^2}\frac{\partial^2 u}{\partial\varphi^2} + \frac{\cot\varphi}{\rho^2}\frac{\partial u}{\partial\varphi} + \frac{1}{\rho^2\sin^2\varphi}\frac{\partial^2 u}{\partial\theta^2}\right) + a(\rho,\varphi,\theta)\dot{u} = 0,\\
&u = 0, \qquad\qquad (t,\rho,\varphi,\theta) \in (0,T)\times\partial\Sigma,\\
&u_0(\rho,\varphi,\theta) = \sin(\pi\rho\sin\varphi\cos\theta)\sin(\pi\rho\sin\varphi\sin\theta)\sin(\pi\rho\cos\varphi),\\
&u_1(\rho,\varphi,\theta) = 0,
\end{aligned}$$

Additionally, we assume that $a(\rho,\varphi,\theta) = 10$ and $L = \frac{1}{10}$. The aim is to obtain the optimal region $\sigma \subset \Sigma$ such that the energy of the system in (1) is minimized at the given end time $T = 1$. We also impose the following additional conditions required by the proposed method:
$u(T=1) = \dot{u}(T=1) = 1$, $(1.6, \frac{\pi}{2}, 0) \in \partial D$, $F_\rho = F_\theta = F_{\rho\theta} = [-5, 5]$, $U = \dot{U} = [0, 1]$, $U_\rho = \left[0, \frac{3}{2}\right]$ and $U_\varphi = U_\theta = [0, 2]$.

For discretization, we consider $N_1 = 3^7 \times 6 \times 5$ number of points in D' and $N_2 = 6 \times 17 \times 8^4$ number of points in D''. To discretize $D^{'} = [0,T] \times [0,2\pi] \times [0,\rho_\Sigma] \times [0,\varphi_\Sigma] \times U \times \dot{U} \times U_\rho \times U_\varphi \times U_\theta$, we choose N_1 points in this set by selecting

1. 3 points in $[0,T]$ for t: $0, \frac{1}{2}, 1$;
2. 3 points in $[0,2\pi]$ for θ: $0, \pi, 2\pi$;
3. 6 points in $[0,2]$ for ρ: $0, \frac{2}{5}, \frac{4}{5}, \frac{6}{5}, \frac{8}{5}, 2$;
4. 5 points in $[0,\varphi_\Sigma]$ for φ: $0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}, \pi$;
5. 3 points in U for u: $0, \frac{1}{2}, 1$;
6. 3 points in $\dot{U}$ for $\dot{u}$: $0, \frac{1}{2}, 1$;
7. 3 points in U_ρ for u_ρ: $0, \frac{3}{4}, \frac{3}{2}$;
8. 3 points in U_φ for u_φ: $0, 1, 2$;
9. 3 points in U_θ for u_θ: $0, 1, 2$;

Also, to discretize $D'' = D_1 \times F_\rho \times F_\theta \times F_{\rho\theta}$, we choose N_2 points in these sets by selecting

1. 6 points in $[0,\rho_{max}]$ for ρ : $0, \frac{2}{5}, \frac{4}{5}, \frac{6}{5}, \frac{8}{5}, 2$;
2. 17 points in $[0,2\pi]$ for θ: $0, \frac{2\pi}{16}, \frac{4\pi}{16}, \frac{6\pi}{16}, \frac{8\pi}{16}, \frac{10\pi}{16}, \frac{12\pi}{16}, \frac{14\pi}{16}, \pi$, $\frac{18\pi}{16}, \frac{20\pi}{16}, \frac{22\pi}{16}, \frac{24\pi}{16}, \frac{26\pi}{16}, \frac{28\pi}{16}, \frac{30\pi}{16}, 2\pi$;
3. 8 values in F_ρ for f_ρ: $-5, \frac{-25}{7}, \frac{-15}{7}, \frac{-5}{7}, \frac{5}{7}, \frac{15}{7}, \frac{25}{7}, 5$;
4. 8 values in F_θ for f_θ: $-5, \frac{-25}{7}, \frac{-15}{7}, \frac{-5}{7}, \frac{5}{7}, \frac{15}{7}, \frac{25}{7}, 5$;
5. 8 values in $F_{\rho\theta}$ for $f_{\rho\theta}$: $-5, \frac{-25}{7}, \frac{-15}{7}, \frac{-5}{7}, \frac{5}{7}, \frac{15}{7}, \frac{25}{7}, 5$.

We then set up the corresponding LP (33) by considering the total functions in the same way as expressed in Section 7 for $M_1 = 3$, $M_2 = 6$, $R_1 = 3$, $R_2 = 6$, $M_3 = 2$, and $M_4 = M_5 = 3$. Therefore, the resulting LPP has $\left(\left(3^7 \times 6 \times 5\right) + \left(6 \times 17 \times 8^4\right)\right)$ variables and 36 constraints.

We use the PSO algorithm to determine the optimal values

$$(\rho_1^*, \ldots, \rho_M^*, \varphi_1^*, \ldots, \varphi_M^*)$$

by implementing the modified simplex method in MATLAB 7.11. After 30 iterations, the optimal points $(\rho_m^*, \varphi_m^*, \theta_m^*)$ of the optimal shape and the optimal value $E^*(\hat{\alpha}, \hat{\beta}, D) = 5.941 \times 10^{-18}$ are obtained.

The obtained solution exhibits a very small objective value of 5.941×10^{-18}. Moreover, the objective value shows no significant improvement after 30 iterations, indicating that the PSO search has stabilized. To further examine the sensitivity of the result to the random initialization, we performed multiple independent runs, all of which converged to the same objective value. These observations provide numerical evidence for the stability and near-optimality of the obtained solution.

The convergence performance of the PSO algorithm is illustrated in Figure 3. As can be seen, the algorithm demonstrates excellent convergence characteristics, achieving a final objective function value of 5.941×10^{-18} after 30 iterations.

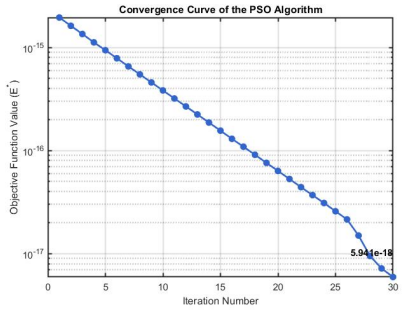


Figure 3: Convergence curve of the PSO algorithm in Example 1.

From Step 3 of the algorithm, the 2D shadow of the unknown surface σ is obtained, as shown by the broken lines in Figure 4. The unknown surface σ is displayed in Figure 5, obtained without imposing the symmetry property. Moreover, the complete 3D damping region is constructed by considering the symmetry property and plotting contour curves of spheres through the optimal points, as illustrated in Figure 6.

A detailed description of the figures for Example 1 is provided in Table 1.

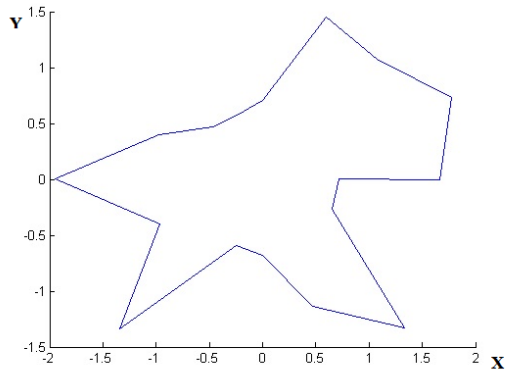


Figure 4: The nearly optimal region D in Example 1.

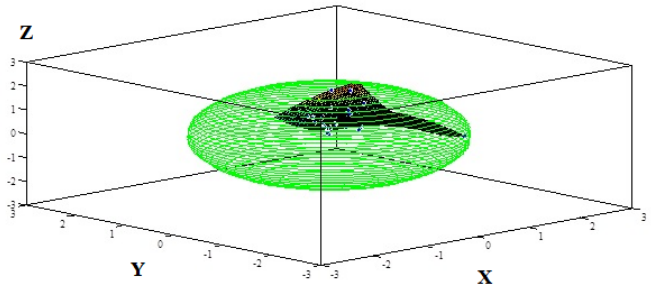


Figure 5: The nearly optimal surface of damping set in the given domain Σ in Example 1.

Example 2. In this example, all conditions and functions are the same as in the previous example, with the difference that the damping coefficient a is also unknown here.

To solve (33), we use an iterative method that consists of two stages. In the first stage, by considering a constant value for the variable a, function $J_1 : [0,1] \longrightarrow \mathbb{R}$ is defined as $J_1(D_M) = E^*(\alpha, \beta, D_M)$. Then, in the second stage, this function is optimized using one of the standard optimization algorithms (like linear search). By doing so, for each value of D_M, a value for the initial objective function is obtained. The set of values obtained for the initial objective function constructs the new objective function. We recall that every time this function is calculated, a linear programming problem similar to (33) with the D_M value identified by this function must be solved.

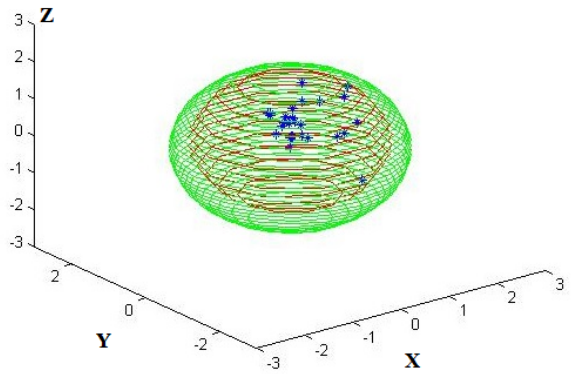


Figure 6: The nearly optimal points that determine the unknown damping set within the domain Σ in Example 1.

Table 1: Detailed description of the figures for Example 1.

Figure	**Description**
Figure 4	The 2D shadow (projection) of the unknown surface that defines the damping region σ, obtained by the shape optimization method. The exact shape of this shadow is not known a priori and is determined by minimizing the energy functional of the 3D wave equation.
Figure 5	The unknown damping region σ inside the spherical region Σ with radius $R = 2$. This figure shows the entire damping region and how it is positioned within the computational domain.
Figure 6	Optimal points in 3D space obtained by a linear programming problem scheme. These points determine the unknown surface σ defining the wave damping region. The surface is visualized using contour lines inside the large sphere Σ to make it clearly observable.

If we denote the optimal value of J_1 as D_M^*, then $J_1^*(D_M) \equiv E^*(\alpha, \beta, a, D_M^*)$. Now, the function $J_2 : [0,1] \to \mathbb{R}$ is calculated as $J_2(a) = J_1^*(D_M^*)$ in the second stage. Therefore, by optimizing the functions J_1 and J_2 and solving the linear programming problem (33), the unknowns a^*, D_M^*, $\hat{\alpha}_1^*, \hat{\alpha}_2^*, \ldots, \hat{\alpha}_N^*$ and $\hat{\beta}_1^*, \hat{\beta}_2^*, \ldots, \hat{\beta}_n^*$ are determined simultaneously.

In this step, we have used the random search algorithm to minimize the function J_1 and the PSO algorithm to optimize the function J_2. The optimal value of $a^* = 24.3624$ and the optimal value of the energy function is 10.323. The optimal region for this example is shown in Figure 7. In [24], for the 2D case and $T = 1$, the optimal value for the variable $a = 29.09$ was found, and

with 2000 iterations using the level set method, the optimal energy value was determined to be 12.56.

As mentioned in this reference, the level set method is entirely dependent on its initial shape, and changing the initial shape will alter the optimal value of the objective function. However, the shape-measure method presented in this paper is independent of the initial shape or value, which is one of the advantages of this method for solving such problems. It is worth noting that in random search and PSO algorithms, the upper and lower bounds for the parameters a and D_M need to be determined. Therefore, when using these algorithms, we considered the unknown a within the range $[1, 50]$ and the unknown D_M based on the shape of the area Σ within the range $[0, 2]$.

The convergence behavior of the proposed algorithm is illustrated in Figure 8. The objective function (E^*) represents the total energy functional that we aim to minimize through optimal design of the damping coefficient distribution.

Figure 7 illustrates the optimal surfaces obtained in the 3D domain. Unlike Figure 6, where only the shape of the damping region was unknown, in this example, the damping coefficient is also an unknown parameter to be identified. The figure displays the optimal surfaces that most efficiently reduce the wave energy to zero. The contour lines represent the level sets of these surfaces within the spherical domain Σ.

The large region Σ is a regular sphere with radius $R = 2$ centered at the origin. The damping region $\sigma \subset \Sigma$ is an unknown subdomain obtained from the optimization process. The goal is to find the optimal shape and coefficient of the damping region such that the total energy of the 3D wave equation decays exponentially to zero.

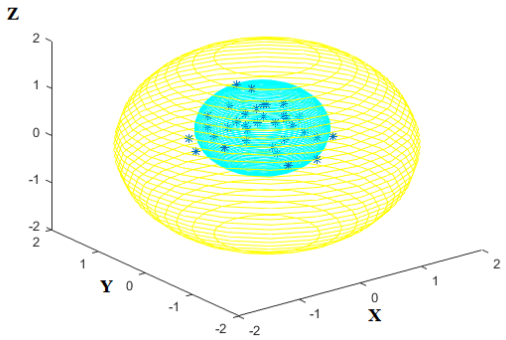


Figure 7: The nearly optimal shape of damping set with an unknown damping coefficient in Example 2.

To study the damping behavior, we draw the wave function $u(t, \rho, \phi, \theta)$ with the optimal damping $a^* = 24.3624$. Figure 9 shows the wave surface at $z = 0.5$, which is equivalent to

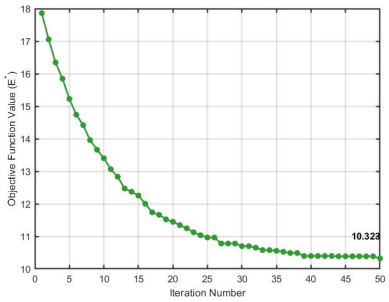


Figure 8: Convergence curve of the PSO algorithm in Example 2.

$\rho cos\phi = 0.5$, at three times: $t = 0$, $t = 0.5$, and $t = 1.0$. The cyan circle on the base shows the damping region. At $t = 0$, the wave has a large peak at the center. At $t = 0.5$, the wave becomes smaller. At $t = 1.0$, the wave is very small; this means that the optimal damping reduces the wave energy well.

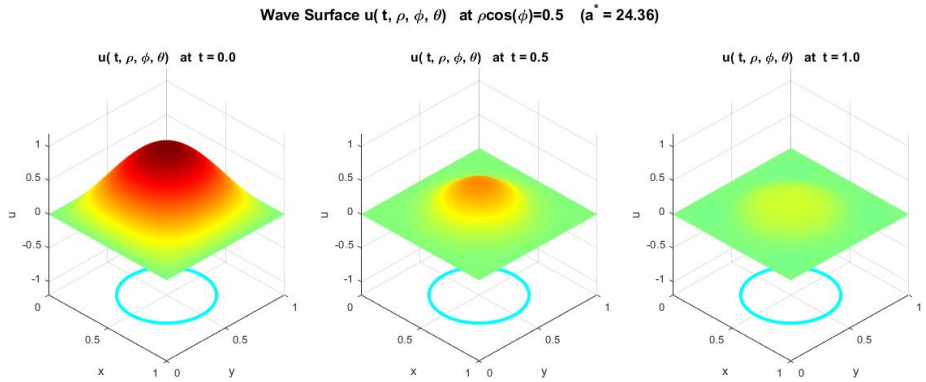


Figure 9: Wave surface $u(t, \rho, \phi, \theta)$ at $\rho cos\phi = 0.5$ for three times: (a) $t = 0$, (b) $t = 0.5$, and (c) $t = 1.0$ in Example 2.

9 Conclusion

We present and apply an extended embedding method to design the shape of a 3D damping set in order to minimize the energy associated with the system of the linear damped wave system. The problem was solved using a two-phase optimization approach based on measure theory. The new method has transformed the classical OSD problem into a nonclassical measure-theoretical problem, which can be formulated as a linear programming problem in measure

space. Besides simplifying the original problem, this method offers several advantages over others: Linear treatment for nonlinear problems, global optimization, time efficiency, and independence from the initial shape and mesh design or additional assumptions on the functions involved in the classical main problem. Therefore, the proposed method can be considered highly suitable for designing the optimal shape in 3D.

As a proposal for future work, the presented method can be extended to solve optimal shape design problems for nonlinear wave equations. Another interesting direction is to apply this method to design optimal damping sets in domains where the wave speed or material properties vary spatially. Furthermore, combining this measure-theoretical approach with intelligent methods, such as neural networks, could help reduce the computational time required to solve complex 3D shape optimization problems.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Alimorad, H. and Fakharzadeh, A. J. *A new approach for solving the obstacle problems in three-dimensions*, Optik, 276 (2023) 170646.

[2] Alimorad, H. and Fakharzadeh, A. J. *Finding the optimal place of sensors for a 3D damped wave equation by using measure approach*, Control Optim. Appl. Math. 4(1) (2019) 15–36.

[3] Alimorad, H. and Fakharzadeh, A. J. *A theoretical measure technique for determining 3D symmetric nearly optimal shapes with a given center of mass*, Comput. Math. & Math. Phys. 57(7) (2017) 1225–1240.

[4] Allaire, G. *Shape Optimization by the Homogenization Method*, Springer, 2007.

[5] Chen, L. L., Lian, H., Liu, Z., Chen, H. B., Atroshchenko, E. and Bordas, S. P. A. *Structural shape optimization of three dimensional acoustic problems with isogeometric boundary element methods*, Comput. Methods Appl. Mech. Eng. 355 (2019) 926–951.

[6] Cheng, D. K. *Field and Wave Electromagnetics*, 2nd ed., Addison-Wesley, 1989.

[7] Fakharzadeh J, A., Alimorad, H. and Tuobaei, S. *A linear technique for designing optimal rotated shapes*, Comput. Appl. Math. 41 (2022) 421.

[8] Fakharzadeh J, A. and Alimorad, H. *A review of theoretical measure approaches in optimal shape problems*, Int. J. Numer. Anal. Model. 16(4) (2019) 543–574.

[9] Fakharzadeh J, A. and Goodarzi, M. *General optimal shape of plasmonic nano particles based absorption coefficient using a linearization approach*, Optik 140 (2017) 1069–1080.

[10] Fakharzadeh J, A., Alimorad, H. and Beiranvand, A. *A linearization technique for optimal design of the damping set with internal dissipation*, Iran. J. Numer. Anal. Optim. 6(1) (2016) 1–29.

[11] Fakharzadeh J, A., Goodarzi, M. and Karimyar, M. J. *On efficiency of some meta-huristic optimization algorithm in shape-measure method*, The first national conference on meta-heuristic algorithms and their applications in science and engineering, Fereydunkenar, 2014.

[12] Fakharzadeh J, A. and Jafarpoor, F. N. *A new discretization method for solving an optimal vibrating controlled system on an arbitrary domain*, J. Vib. Control 19(14) (2012) 2092–2099.

[13] Fakharzadeh J, A. and Rubio, J. E. *Global solution of optimal shapes design problems*, J. Anal. Appl. 18(1) (1999) 143–155.

[14] Farahi, M. H., Borzabadi, A. H., Mehneh, H. H. and Kamyad, A. V. *Measure theoretical approach for optimal shape design of a nozzle*, J. Appl. Math. Comput. 17(1-2) (2005) 315–328.

[15] Henrot, A. and Pierre, M. *Shape Variation and Optimization: A Geometrical Analysis*, EMS Publishing House, 2018.

[16] Hawkins, D. *Identification of Outliers*, Chapman and Hall, London, 1980.

[17] Hébrard, P. and Henrot, A. *Optimal shape and position of the actuators for the stabilization of a string*, Syst. Control Lett. 48(3-4) (2003) 199–209.

[18] Hesse, S. H., Leidinger, L. F., Kremheller, J., Lukaszewicz, D. H. J. A. and Duddeck, F. *Shape optimization with the level-set-method using local volume constraints*, Struct. Multidiscip. Optim. 57(1) (2018) 115–130.

[19] Kriegel, H.-P., Kroger, P., Schubert, E. and Zimek, A. *LoOP: Local outlier probabilities*, In Proc ACM Conference on Information and Knowledge Management (CIKM), Hong Kong, China, 2009, 2–6.

[20] Liu, M. and Hasegawa, Y. *Adjoint-based shape optimization for compressible flow based on volume penalization method*, Eng. Comput. 41 (2025) 989–1010.

[21] Lu, C., Liu, J. and Guo, A. *Three-dimensional shape optimization of a submerged body under wave diffraction*, Phys. Fluids 35(10) (2023) 107116.

[22] Mehne, H. H. and Farahi, M. H. *Transformation to a fixed domain in LP modelling for a class of optimal shape design problems*, Iran. J. Numer. Anal. Optim. 9(1) (2019) 1–16.

[23] Mikhailov, V. P. *Partial Differential Equations*, MIR Publisher, Moscow, 1978.

[24] Münch, A. *Optimal internal dissipation of a damped wave equation using a topological approach*, Int. J. Appl. Math. Comput. Sci. 19(1) (2009) 15–37.

[25] Münch, A., Pedregal, P. and Periago, F. *Optimal design of the damping set for the stabilization of the wave equation*, J. Differ. Equ. 231(1) (2006) 331–358.

[26] Münch, A., Pedregal, P. and Periago, F. *Optimal internal stabilization of the linear system of elasticity*, Arch. Ration. Mech. Anal. 193(1) (2009) 171–193.

[27] Pironneau, O. *Optimal Shape Design for Elliptic Systems*, Springer, New York, 1983.

[28] Rubio, J. E. *Control and Optimization: The Linear Treatment of Nonlinear Problems*, Manchester University Press, Manchester, U.K., 1986.

[29] Rudin, W. *Real and Complex Analysis*, 2nd edn., Tata McGraw-Hill Publishing Co Ltd., New Delhi, 1983.

[30] Takahashi, T., Sato, D., Isakari, H. and Matsumoto, T. *A shape optimisation with the isogeometric boundary element method and adjoint variable method for the three-dimensional Helmholtz equation*, Comput.-Aided Des. 143 (2022) 103126.

[31] Tcheugoué Tébou, L. R. *Stabilization of the wave equation with localized nonlinear damping*, J. Differ. Equ. 145 (1998) 502–524.

[32] Thomas George, B. and Finney Ross, L. *Calculus and Analytic Geometry*, 9th edn., Addison-Wesley, United States of America, 1998.

[33] Vineyard, E. and Gao, X. L. *Topology and shape optimization of 2D and 3D micro-architectured thermoelastic metamaterials using a parametric level set method*, CMES-Comput. Model. Eng. Sci. 127(3) (2021) 819–854.

[34] Wang, K., Zhang, X. Y., Zhang, Z. D. and Min, C. H. *Three-dimensional shape optimization of fins in a printed circuit recuperator using S-CO2 as the heat-transfer fluid*, Int. J. Heat Mass Transf. 192 (2022) 122910.

[35] Wang, K., Liu, Y. J., Zhang, Z. D., Zhang, X., Fan, Y. H. and Min, C. H. *Three-dimensional shape optimization of fins for application in compact supercritical CO2 solar receivers*, Int. J. Heat Mass Transf. 221 (2024) 125013.

[36] Zuazua, E. *Propagation, observation, and control of waves approximated by finite difference methods*, SIAM Rev. 47(2) (2005) 197–243.

Iranian Journal of Numerical Analysis and Optimization
Vol. 17, No. 1, 2027, pp 494–528
https://doi.org/10.22067/ijnao.2026.99100.1885
https://ijnao.um.ac.ir/

CC BY NC

Research Article

Stability and sensitivity analysis of a nonlinear compartmental model for banking risk contagion

S. Ibnoukady*, A. Meskaf and K. Allali

Abstract

In this work, we study a mathematical model that describes the dynamics of a nonlinear financial system and explains how systemic risk propagates within a banking system. The proposed model is formulated as a system of differential equations structured into five compartments: Unexposed, exposed, distressed, recovered, and liquidated. Two mechanisms are incorporated: A direct early-intervention pathway that allows distressed banks to return directly to the unexposed state and a reintegration channel from the liquidated state to the unexposed state. Using the next-generation matrix approach, we derive the risk reproduction number and establish the system's stability properties. We analyze the equilibrium points and prove the global asymptotic stability of the risk-free equilibrium via a Lyapunov function and the local and global asymptotic stability of the persistent-risk equilibrium via the Routh–Hurwitz criterion

*Corresponding author

Received 3 June 2026; revised 13 August 2026; accepted 28 September 2026

Soukaina Ibnoukady
Experimental Laboratory of Innovation in Technology and Simulations, Faculty of Sciences, Chouaib Doukkali University, El Jadida, Morocco. e-mail: soukaina.i419@ucd.ac.ma

Adil Meskaf
Department of SEG, Faculty of Economic and Social Legal Sciences, Chouaib Doukkali University, El Jadida, Morocco. e-mail: meskaf.a@ucd.ac.ma

Karam Allali
Laboratory of Mathematics, Computer Science and Applications, Faculty of Sciences and Technologies, Hassan II University of Casablanca, Mohammedia, Morocco. e-mail: allali@hotmail.com

How to cite this article
Ibnoukady, S., Meskaf, A. and Allali, K., Stability and sensitivity analysis of a nonlinear compartmental model for banking risk contagion. *Iran. J. Numer. Anal. Optim.*, 2027; 17(1): 494-528. https://doi.org/10.22067/ijnao.2026.99100.1885

and a Lyapunov function. Finally, the numerical simulations and the LHS-PRCC sensitivity analysis illustrate the effects of contagion, intervention, and reintegration on the dynamics of the system.

AMS subject classifications (2020): Primary 91G10; Secondary 91B30, 34D20, 34D23.

Keywords: Systemic risk; Banking contagion; Mathematical modelling; Stability analysis; Risk reproduction number.

1 Introduction

The financial system is highly interconnected and remains fragile. It is therefore important to study how systemic risk spreads in order to preserve overall stability [29, 30]. Past crises, such as the collapse of Lehman Brothers in 2008, have shown that a local problem can trigger an international chain reaction, disrupt credit markets, and lead to a sharp economic slowdown [4, 25]. The failure of Silicon Valley Bank in 2023 also showed that this problem still exists despite the reforms introduced after previous crises [18]. These events show the vulnerability of banking systems, where financial difficulties can spread through the links and contracts between banks [1]. For this reason, we need reliable tools to describe the evolution of financial contagion [10].

In recent years, mathematical models have helped us understand how risk spreads between banks [30, 2, 17, 26]. These studies show that because bank balance sheets are linked, the failure of one institution can weaken the whole financial system. However, some questions remain about the effects of intervention measures and the ability of banks to recover after a shock [19, 5]. It is therefore useful to develop models that describe these mechanisms more clearly [24, 8].

Systemic risk has received growing attention in recent financial studies. Financial network models explain how financial shocks spread through interbank links and central clearing [10, 29]. Banks are also linked through their balance sheets. Losses can spread from one bank to another through direct exposures, common asset holdings, and fire sales [15]. Macroprudential policies can help reduce systemic risk. However, their effects depend on the measure used and the characteristics of banks [28]. In this work, the proposed compartmental model describes how banks move from one financial state to another. It also studies the effects of early intervention and reintegration.

In this work, we introduce a nonlinear compartmental model for the spread of systemic risk in a banking system. Banks are divided into five compartments: Unexposed (U), Exposed (E), Distressed (D), Recovered (R), and Liquidated (L). The models of Irakoze et al. [16] and Ait

Tamerz, Allali, and Meskaf [2] are also based on these five UEDRL compartments. The first study uses an integer-order system, while the second uses a fractional-order system. Table 1 compares the main contributions of these models and the present study.

Table 1: Comparison of the UEDRL models.

Study	Model	Mathematical contribution	Financial contribution
Irakoze et al. (2023) [16]	Integer-order UEDRL model.	Analysis of the equilibria, their stability, and numerical simulations.	Describes the spread and persistence of systemic risk in a banking system.
Ait Tamerz, Allali, and Meskaf (2024) [2]	Fractional-order UEDRL model.	Extends the UEDRL model using the Caputo derivative and studies its stability.	Takes into account the effect of past financial conditions on systemic risk.
Present study	Integer-order UEDRL model with the transitions $D \to U$ and $L \to U$.	Studies the effects of early intervention and reintegration on the risk reproduction number, the equilibria, and system stability.	Represents early intervention and the reintegration of viable banking activities.

The mathematical tools used in this study are standard, but their application depends on the specific structure of our model. We first prove the existence, positivity, and boundedness of the solutions and calculate the risk reproduction number R_0 using the next-generation matrix method [35]. Since the total number of banks is conserved, the Jacobian matrix of the full system has a zero eigenvalue. We therefore remove one variable and study stability on a four-dimensional invariant set. At the persistent-risk equilibrium, the reduced system gives a fourth-degree characteristic polynomial, and the Routh–Hurwitz conditions are verified using the equilibrium relations. We also study the equilibrium points using Lyapunov functions and LaSalle's invariance principle, where the Lyapunov functions are chosen according to the transfers between the compartments and the return paths to U [22, 31, 34]. Finally, numerical simulations are used to study the effects of early intervention and reintegration on the dynamics of the banking system.

The paper is organized as follows: Section 2 provides the formulation of the model and parameters. Section 3 describes the basic properties, including positivity and boundedness of solutions. Section 4 defines the risk reproduction number R_0 in order to determine whether the crisis disappears or persists. Section 5 focuses on the stability analysis of the financial equilibria. Numerical simulations are presented in Section 6. Section 7 provides the sensitivity analysis. Finally, Section 8 concludes the paper.

2 Mathematical model formulation

This section introduces the mathematical framework used to represent the propagation of financial distress within an interconnected banking system. Institutions are divided into compartments according to their financial condition, which makes it possible to follow their transition from one state to another over time.

According to these assumptions, the evolution of the banking system is described by the following nonlinear system:

$$\begin{cases} \dfrac{dU}{dt} = -\beta UD + \mu E + \theta R + \delta_3 D + \eta L, \\ \dfrac{dE}{dt} = \beta UD - (\mu + \gamma)E, \\ \dfrac{dD}{dt} = \gamma E - (\delta_1 + \delta_2 + \delta_3)D, \\ \dfrac{dR}{dt} = \delta_1 D - \theta R, \\ \dfrac{dL}{dt} = \delta_2 D - \eta L. \end{cases} \tag{1}$$

In system (1), exposed banks return to the unexposed class at rate μ or become distressed at rate γ. Distressed banks move to the recovered class at rate δ_1 or to the liquidated class at rate δ_2. Recovered banks return to the unexposed class at rate θ. The parameters δ_3 and η represent the direct intervention and reintegration mechanisms, respectively. The term βUD represents a mass-action contagion mechanism. From an economic perspective, distressed banks can transmit losses to initially unexposed banks through interbank exposures, common asset holdings, and fire sales [8, 1, 15]. The product UD approximates the number of potential interactions between these two groups, while β represents their average intensity and the probability that such an interaction produces a new exposure. This formulation is therefore suitable as a first approximation for the closed and homogeneous banking system considered here. The compartments U, E, D, R, and L represent unexposed, exposed, distressed, recovered, and liquidated banks, respectively. Liquidated banks may return after acquisition or reorganization [16].

The interactions between these compartments are illustrated in Figure 1. The diagram highlights the main distress and recovery pathways, including the direct intervention mechanism δ_3 and the reintegration mechanism η.

All parameters are assumed to be strictly positive. Table 2 summarises their economic and mathematical roles [13, 30].

The parameters δ_3 and η represent two mechanisms used in banking crisis management. The parameter δ_3 represents early intervention, such as recapitalization, debt restructuring, or

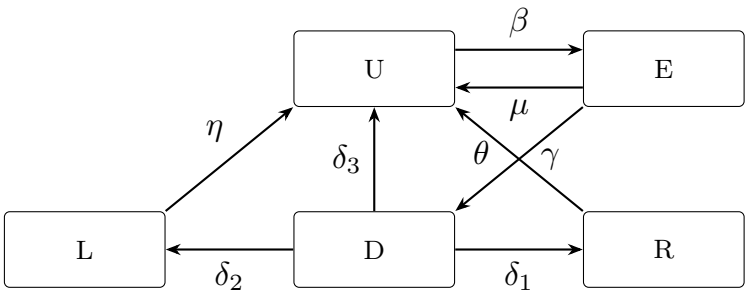


Figure 1: Compartmental structure of the UEDRL banking contagion model.

Table 2: Interpretation of the model parameters.

Parameter	**Interpretation**
β	Contagion intensity between unexposed and distressed banks, associated with the nonlinear term βUD.
μ	Recovery rate of exposed banks returning directly to the unexposed class.
γ	Deterioration rate of exposed banks moving to the distressed class.
δ_1	Restructuring or resolution rate of distressed banks moving to the recovered class.
δ_2	Liquidation rate of distressed banks temporarily leaving the active system.
δ_3	Direct intervention rate allowing distressed banks to return immediately to the unexposed class.
θ	Reintegration rate of recovered banks into the unexposed class.
η	Reintegration rate of viable banking activities after acquisition or reorganization

a change in management, which may help a distressed bank return from D to U [3]. The parameter η represents the transfer of viable activities from a failed bank to another bank or a bridge bank [32]. Thus, the transition from L to U represents the reintegration of these activities, not the return of the same liquidated bank.

The two new parameters do not have the same mathematical effect. The parameter δ_3 appears in the risk reproduction number because it increases the exit rate from the distressed compartment. The parameter η does not appear in this number, but it makes the liquidated compartment nonterminal and changes the persistent-risk equilibrium.

We assume a closed banking system where the total number of institutions remains constant:

$$N(t) = U(t) + E(t) + D(t) + R(t) + L(t) = N_0. \tag{2}$$

Indeed, by summing the five equations of system (1), we obtain

$$\frac{dN}{dt} = \frac{dU}{dt} + \frac{dE}{dt} + \frac{dD}{dt} + \frac{dR}{dt} + \frac{dL}{dt} = 0. \tag{3}$$

Thus, $N(t) = N(0) = N_0$, which confirms that the total number of banks remains constant.

Finally, defining the state vector

$$X(t) = \big(U(t), E(t), D(t), R(t), L(t)\big)^\top,$$

system (1) can be written compactly as

$$\frac{dX}{dt} = F(X),$$

where F is a nonlinear vector field. Since the right-hand side of the system is polynomial in the state variables, it is locally Lipschitz. This property ensures the local existence and uniqueness of solutions and provides the basis for the qualitative analysis developed in the following section [35].

The deterministic model describes the average movement of banks and is used to calculate the risk reproduction number. It assumes that banks are similar and interact in the same way. The term βUD represents a continuous propagation of risk. However, it does not represent sudden changes caused by large losses, fire sales, or the failure of a highly connected bank. Differences in bank size, network structure, and delays in risk transmission are also not included. These aspects may be considered in future studies by using other incidence functions, banking networks, or sudden changes in the model parameters.

3 Basic properties of the model

In this section, we establish the main qualitative properties of system (1). These properties are important because all state variables represent numbers of banks and must therefore remain nonnegative and bounded for all time [33, 34].

Let

$$X(t) = \big(U(t), E(t), D(t), R(t), L(t)\big)^\top \in \mathbb{R}^5_+$$

be the solution of system (1). All parameters are assumed to be strictly positive.

3.1 Local existence and uniqueness

Let $F : \mathbb{R}^5 \to \mathbb{R}^5$ be the vector field defined by the right-hand side of system (1). Since each component of F is a polynomial function of (U, E, D, R, L), we have

$$F \in C^\infty(\mathbb{R}^5).$$

Hence, F is locally Lipschitz on $\mathbb{R}^5$. By the classical Cauchy–Lipschitz theorem [7, 21], for any nonnegative initial condition

$$X(0) = \big(U(0), E(0), D(0), R(0), L(0)\big)^\top \in \mathbb{R}^5_+,$$

there exists a time $T > 0$ and a unique local solution of system (1) on $[0, T)$.

3.2 Positivity of solutions

Since the variables describe banking institutions, it is necessary to show that the nonnegative orthant is positively invariant.

Proposition 1. If $X(0) \in \mathbb{R}^5_+$, then the solution of system (1) satisfies

$$X(t) \in \mathbb{R}^5_+, \qquad \text{for all } t \geq 0.$$

Proof. We verify the sign of the vector field on the boundary of the nonnegative orthant. On each boundary face, we obtain

$$\begin{cases} \left.\dfrac{dU}{dt}\right|_{U=0} = \mu E + \theta R + \delta_3 D + \eta L \geq 0, \\ \left.\dfrac{dE}{dt}\right|_{E=0} = \beta U D \geq 0, \\ \left.\dfrac{dD}{dt}\right|_{D=0} = \gamma E \geq 0, \\ \left.\dfrac{dR}{dt}\right|_{R=0} = \delta_1 D \geq 0, \\ \left.\dfrac{dL}{dt}\right|_{L=0} = \delta_2 D \geq 0. \end{cases}$$

Thus, the vector field points inward or is tangent to the boundary of $\mathbb{R}^5_+$. Therefore, no solution starting from a nonnegative initial condition can leave the nonnegative orthant. Hence, $\mathbb{R}^5_+$ is positively invariant [27, 34]. □

3.3 Conservation and boundedness

From the conservation relation established above, we have

$$N(t) = N(0) = N_0, \qquad \text{for all } t \geq 0. \tag{4}$$

This conservation property is standard in closed compartmental systems and is used to define a bounded feasible region for the model [34, 21, 14]. Since all components are nonnegative, it follows that

$$0 \leq U(t), E(t), D(t), R(t), L(t) \leq N_0, \qquad \text{for all } t \geq 0. \tag{5}$$

Thus, all solutions remain bounded for all time.

3.4 Invariant feasible region

From the positivity and conservation properties, the natural feasible region of the model is

$$\Omega = \left\{(U, E, D, R, L) \in \mathbb{R}^5_+ : U + E + D + R + L = N_0\right\}. \tag{6}$$

If $X(0) \in \Omega$, then positivity gives $X(t) \in \mathbb{R}^5_+$, and the conservation law (4) gives

$$U(t) + E(t) + D(t) + R(t) + L(t) = N_0.$$

Hence,

$$X(t) \in \Omega, \qquad \text{for all } t \geq 0.$$

Therefore, Ω is positively invariant for system (1). This follows from the standard invariance arguments used for positive dynamical systems on closed and bounded feasible regions [33, 34].

3.5 Global existence

Since Ω is closed and bounded in $\mathbb{R}^5$, it is compact. Moreover, the vector field F is locally Lipschitz and all solutions starting in Ω remain bounded in Ω. Hence, no finite-time blow-up can occur, and the local solution can be extended to all $t \geq 0$ [7, 34].

Proposition 2. For every initial condition $X(0) \in \Omega$, system (1) admits a unique global solution satisfying

$$X(t) \in \Omega, \qquad \text{for all } t \geq 0.$$

4 Equilibria and risk reproduction number

In this section, we determine the equilibrium points of system (1) and derive the risk reproduction number R_0. This quantity plays an important role in the analysis of financial contagion, since it allows us to distinguish between the disappearance and the persistence of systemic distress.

4.1 Risk-free equilibrium

The risk-free equilibrium corresponds to the situation where no bank is exposed or distressed. In this case,

$$E = 0, \qquad D = 0.$$

At equilibrium, from the equations of R and L, we obtain

$$-\theta R = 0, \qquad -\eta L = 0.$$

Since $\theta > 0$ and $\eta > 0$, it follows that

$$R = 0, \qquad L = 0.$$

Using the conservation law

$$U + E + D + R + L = N_0,$$

we get $U = N_0$. Therefore, the system admits the following risk-free equilibrium:

$$X_0 = (N_0, 0, 0, 0, 0). \tag{7}$$

4.2 Persistent-risk equilibrium

A persistent-risk equilibrium is an equilibrium point at which financial distress remains present in the system. We denote it by

$$X^* = (U^*, E^*, D^*, R^*, L^*),$$

with

$$E^* > 0, \qquad D^* > 0.$$

At equilibrium, system (1) gives

$$0 = -\beta U^* D^* + \mu E^* + \theta R^* + \delta_3 D^* + \eta L^*, \tag{8}$$

$$0 = \beta U^* D^* - (\mu + \gamma) E^*, \tag{9}$$

$$0 = \gamma E^* - (\delta_1 + \delta_2 + \delta_3) D^*, \tag{10}$$

$$0 = \delta_1 D^* - \theta R^*, \tag{11}$$

$$0 = \delta_2 D^* - \eta L^*. \tag{12}$$

From (10), (11), and (12), we obtain

$$E^* = \frac{\delta_1 + \delta_2 + \delta_3}{\gamma} D^*, \qquad R^* = \frac{\delta_1}{\theta} D^*, \qquad L^* = \frac{\delta_2}{\eta} D^*. \tag{13}$$

Moreover, from (9), since $D^* > 0$, we get

$$U^* = \frac{(\mu + \gamma)(\delta_1 + \delta_2 + \delta_3)}{\beta \gamma}. \tag{14}$$

Using the conservation law, we have

$$U^* + E^* + D^* + R^* + L^* = N_0.$$

Substituting (13) gives

$$D^* \left(1 + \frac{\delta_1 + \delta_2 + \delta_3}{\gamma} + \frac{\delta_1}{\theta} + \frac{\delta_2}{\eta} \right) = N_0 - U^*. \tag{15}$$

Hence,

$$D^* > 0 \quad \Longleftrightarrow \quad N_0 > U^*.$$

Using (14), this condition becomes

$$\frac{\beta N_0 \gamma}{(\mu+\gamma)(\delta_1+\delta_2+\delta_3)} > 1. \tag{16}$$

This motivates the definition of the risk reproduction number R_0 [9, 35].

4.3 Derivation of the risk reproduction number

To derive R_0, we use the next-generation matrix approach applied to the exposed and distressed compartments [9, 35]. Around the risk-free equilibrium $X_0 = (N_0, 0, 0, 0, 0)$, the subsystem associated with E and D can be written as

$$\frac{d}{dt}\begin{pmatrix} E \\ D \end{pmatrix} = \mathcal{F}(E,D) - \mathcal{V}(E,D),$$

where

$$\mathcal{F}(E,D) = \begin{pmatrix} \beta N_0 D \\ 0 \end{pmatrix}, \qquad \mathcal{V}(E,D) = \begin{pmatrix} (\mu+\gamma)E \\ -\gamma E + (\delta_1+\delta_2+\delta_3)D \end{pmatrix}. \tag{17}$$

The corresponding Jacobian matrices at X_0 are

$$F = \begin{pmatrix} 0 & \beta N_0 \\ 0 & 0 \end{pmatrix}, \qquad V = \begin{pmatrix} \mu+\gamma & 0 \\ -\gamma & \delta_1+\delta_2+\delta_3 \end{pmatrix}. \tag{18}$$

Since

$$\det(V) = (\mu+\gamma)(\delta_1+\delta_2+\delta_3) > 0,$$

the matrix V is invertible and

$$V^{-1} = \frac{1}{(\mu+\gamma)(\delta_1+\delta_2+\delta_3)} \begin{pmatrix} \delta_1+\delta_2+\delta_3 & 0 \\ \gamma & \mu+\gamma \end{pmatrix}. \tag{19}$$

Thus, the next-generation matrix is

$$K = FV^{-1} = \frac{\beta N_0}{(\mu+\gamma)(\delta_1+\delta_2+\delta_3)} \begin{pmatrix} \gamma & \mu+\gamma \\ 0 & 0 \end{pmatrix}. \tag{20}$$

The risk reproduction number is defined as the spectral radius of K:

$$R_0 = \rho(K). \tag{21}$$

The eigenvalues of K are

$$\lambda_1 = \frac{\beta N_0 \gamma}{(\mu+\gamma)(\delta_1+\delta_2+\delta_3)}, \qquad \lambda_2 = 0.$$

Therefore,

$$R_0 = \frac{\beta N_0 \gamma}{(\mu+\gamma)(\delta_1+\delta_2+\delta_3)}. \tag{22}$$

4.4 Existence of the persistent-risk equilibrium

From (16) and (22), the persistent-risk equilibrium exists if and only if

$$R_0 > 1.$$

More precisely, when $R_0 > 1$, the equilibrium is given by

$$X^* = (U^*, E^*, D^*, R^*, L^*), \tag{23}$$

where

$$U^* = \frac{(\mu+\gamma)(\delta_1+\delta_2+\delta_3)}{\beta\gamma}, \tag{24}$$

$$D^* = \frac{N_0 - U^*}{1 + \dfrac{\delta_1+\delta_2+\delta_3}{\gamma} + \dfrac{\delta_1}{\theta} + \dfrac{\delta_2}{\eta}}, \tag{25}$$

and

$$E^* = \frac{\delta_1+\delta_2+\delta_3}{\gamma} D^*, \qquad R^* = \frac{\delta_1}{\theta} D^*, \qquad L^* = \frac{\delta_2}{\eta} D^*. \tag{26}$$

Thus, the model has a unique risk-free equilibrium for all parameter values, and a unique persistent-risk equilibrium whenever $R_0 > 1$.

Economically, the condition $R_0 < 1$ means that the propagation of financial distress is not strong enough to persist in the network. In contrast, when $R_0 > 1$, distress can remain in the system and the banking network converges toward a state in which some institutions remain exposed or distressed.

5 Stability analysis

In this section, we study the stability properties of the equilibria obtained in Section 4 [20, 22, 31, 34]. We first analyse the local stability of the risk-free equilibrium. Then, we examine the persistent-risk equilibrium when $R_0 > 1$.

5.1 Local stability of the risk-free equilibrium

The risk-free equilibrium is given by

$$X_0 = (N_0, 0, 0, 0, 0).$$

The Jacobian matrix associated with system (1) is

$$J(X) = \begin{pmatrix} -\beta D & \mu & -\beta U + \delta_3 & \theta & \eta \\ \beta D & -(\mu+\gamma) & \beta U & 0 & 0 \\ 0 & \gamma & -(\delta_1+\delta_2+\delta_3) & 0 & 0 \\ 0 & 0 & \delta_1 & -\theta & 0 \\ 0 & 0 & \delta_2 & 0 & -\eta \end{pmatrix}.$$

Evaluating this matrix at X_0, we obtain

$$J(X_0) = \begin{pmatrix} 0 & \mu & -\beta N_0 + \delta_3 & \theta & \eta \\ 0 & -(\mu+\gamma) & \beta N_0 & 0 & 0 \\ 0 & \gamma & -(\delta_1+\delta_2+\delta_3) & 0 & 0 \\ 0 & 0 & \delta_1 & -\theta & 0 \\ 0 & 0 & \delta_2 & 0 & -\eta \end{pmatrix}. \tag{27}$$

Since the total population is conserved, the stability on the feasible region Ω is governed by the exposed-distressed subsystem. The corresponding reduced matrix is

$$A = \begin{pmatrix} -(\mu+\gamma) & \beta N_0 \\ \gamma & -(\delta_1+\delta_2+\delta_3) \end{pmatrix}. \tag{28}$$

Its characteristic polynomial is

$$P(\lambda) = \lambda^2 + a_1\lambda + a_0, \tag{29}$$

where

$$a_1 = (\mu+\gamma) + (\delta_1+\delta_2+\delta_3), \tag{30}$$

and

$$a_0 = (\mu+\gamma)(\delta_1+\delta_2+\delta_3) - \beta N_0\gamma. \tag{31}$$

Since all parameters are strictly positive, we have $a_1 > 0$. Moreover,

$$a_0 > 0 \quad \Longleftrightarrow \quad \frac{\beta N_0\gamma}{(\mu+\gamma)(\delta_1+\delta_2+\delta_3)} < 1.$$

Using the expression of the risk reproduction number R_0, this condition becomes

$$a_0 > 0 \quad \Longleftrightarrow \quad R_0 < 1.$$

By the Routh–Hurwitz criterion for second-order polynomials [20], all roots of $P(\lambda)$ have negative real parts if and only if

$$a_1 > 0 \qquad \text{and} \qquad a_0 > 0.$$

Hence, the risk-free equilibrium is locally asymptotically stable when $R_0 < 1$.

If $R_0 > 1$, then $a_0 < 0$, and the matrix A has one positive eigenvalue. Therefore, the risk-free equilibrium becomes unstable.

Theorem 1. The risk-free equilibrium $X_0 = (N_0, 0, 0, 0, 0)$ of system (1) is locally asymptotically stable in Ω if $R_0 < 1$, and unstable if $R_0 > 1$.

Proof. The local dynamics near X_0 is determined by the linearized exposed-distressed subsystem (28). The characteristic polynomial of this subsystem is given by (29). Since $a_1 > 0$ and $a_0 > 0$ if and only if $R_0 < 1$, the Routh–Hurwitz criterion implies that all eigenvalues of the reduced subsystem have negative real parts. Therefore, the risk-free equilibrium is locally

asymptotically stable when $R_0 < 1$. If $R_0 > 1$, then $a_0 < 0$, which implies that one eigenvalue is positive. Hence, X_0 is unstable. □

The previous result has a direct economic interpretation. When $R_0 < 1$, a small initial financial disturbance cannot spread through the network and the system returns to the risk-free state. When $R_0 > 1$, the initial disturbance is amplified, and the banking system may move toward a persistent-risk regime.

5.2 Local stability of the persistent-risk equilibrium

Assume that $R_0 > 1$. Then the persistent-risk equilibrium

$$X^* = (U^*, E^*, D^*, R^*, L^*)$$

exists in Ω and satisfies $E^* > 0$ and $D^* > 0$. Define

$$\alpha = \mu + \gamma, \qquad \Delta = \delta_1 + \delta_2 + \delta_3, \qquad q = \beta D^* > 0.$$

At equilibrium, we have

$$\beta\gamma U^* = \alpha\Delta. \tag{32}$$

Since the total number of banks is conserved, we use

$$U = N_0 - E - D - R - L$$

and study the reduced system in

$$Y = (E, D, R, L)^\top.$$

The Jacobian matrix of the reduced system at $Y^* = (E^*, D^*, R^*, L^*)^\top$ is

$$J_4(Y^*) = \begin{pmatrix} -(\alpha+q) & \dfrac{\alpha\Delta}{\gamma} - q & -q & -q \\ \gamma & -\Delta & 0 & 0 \\ 0 & \delta_1 & -\theta & 0 \\ 0 & \delta_2 & 0 & -\eta \end{pmatrix}. \tag{33}$$

Its characteristic polynomial is

$$p(\lambda) = \lambda^4 + a_1\lambda^3 + a_2\lambda^2 + a_3\lambda + a_4, \tag{34}$$

where

$$\begin{aligned}
a_1 &= \alpha + q + \Delta + \theta + \eta,\\
a_2 &= \Delta\eta + \Delta q + \Delta\theta + \alpha\eta + \alpha\theta + \eta q + \eta\theta + \gamma q + q\theta,\\
a_3 &= \Delta\eta q + \Delta\eta\theta + \Delta q\theta + \alpha\eta\theta + \delta_1\gamma q + \delta_2\gamma q + \eta\gamma q + \eta q\theta + \gamma q\theta,\\
a_4 &= \Delta\eta q\theta + \delta_1\eta\gamma q + \delta_2\gamma q\theta + \eta\gamma q\theta.
\end{aligned}$$

Theorem 2. If $R_0 > 1$, then the persistent-risk equilibrium X^* is locally asymptotically stable in Ω.

Proof. Since all parameters are strictly positive and $q > 0$, we have

$$a_1 > 0, \qquad a_2 > 0, \qquad a_3 > 0, \qquad a_4 > 0.$$

For the fourth-order polynomial (34), the remaining Routh–Hurwitz conditions are

$$a_1a_2 - a_3 > 0 \tag{35}$$

and

$$a_1a_2a_3 - a_3^2 - a_1^2a_4 > 0. \tag{36}$$

The complete algebraic verification of (35)–(36) is provided in Appendix A. Hence, all eigenvalues of $J_4(Y^*)$ have negative real parts. By the linearization theorem [20], Y^* is locally asymptotically stable for the reduced system. The conservation relation then implies that X^* is locally asymptotically stable in Ω. □

From an economic point of view, in the persistent-risk regime, small changes around the equilibrium remain limited. Under the Routh–Hurwitz conditions, the system gradually returns to the same long-run risk level.

5.3 Global stability of the risk-free equilibrium

We now prove the global asymptotic stability of the risk-free equilibrium in the feasible region Ω when $R_0 < 1$ [22, 31, 34].

Theorem 3. If $R_0 < 1$, then the risk-free equilibrium

$$X_0 = (N_0, 0, 0, 0, 0)$$

is globally asymptotically stable in Ω.

Proof. Consider the Lyapunov function

$$V(E, D) = E + \frac{\mu + \gamma}{\gamma} D. \tag{37}$$

Clearly, $V(E, D) \geq 0$ for all $(E, D) \in \mathbb{R}^2_+$, and

$$V(E, D) = 0 \quad \Longleftrightarrow \quad E = D = 0.$$

Differentiating V along the solutions of system (1), we obtain

$$\dot{V} = \dot{E} + \frac{\mu + \gamma}{\gamma} \dot{D}.$$

Using the equations of E and D, this gives

$$\dot{V} = \beta U D - (\mu + \gamma)E + \frac{\mu + \gamma}{\gamma} \left(\gamma E - (\delta_1 + \delta_2 + \delta_3)D\right).$$

Hence,

$$\dot{V} = D \left[\beta U - \frac{(\mu + \gamma)(\delta_1 + \delta_2 + \delta_3)}{\gamma}\right]. \tag{38}$$

Since $X(t) \in \Omega$, we have $U(t) \leq N_0$. Therefore,

$$\dot{V} \leq D \left[\beta N_0 - \frac{(\mu + \gamma)(\delta_1 + \delta_2 + \delta_3)}{\gamma}\right].$$

Using the definition of R_0, we can write

$$\beta N_0 = R_0 \frac{(\mu + \gamma)(\delta_1 + \delta_2 + \delta_3)}{\gamma}.$$

Thus,

$$\dot{V} \leq \frac{(\mu+\gamma)(\delta_1+\delta_2+\delta_3)}{\gamma}(R_0-1)D. \tag{39}$$

If $R_0 < 1$, then $\dot{V} \leq 0$ in Ω, and $\dot{V} = 0$ implies $D = 0$. From the equation of D,

$$\dot{D} = \gamma E - (\delta_1+\delta_2+\delta_3)D,$$

the condition $D = 0$ gives $\dot{D} = \gamma E$. For the set $\{D = 0\}$ to be invariant, we must also have $E = 0$.

On the largest invariant subset of $\{\dot{V} = 0\}$, we therefore have

$$E = D = 0.$$

Then the equations of R and L reduce to

$$\dot{R} = -\theta R, \qquad \dot{L} = -\eta L,$$

which imply

$$R(t) \to 0, \qquad L(t) \to 0 \quad \text{as } t \to \infty.$$

Finally, using the conservation law

$$U + E + D + R + L = N_0,$$

we obtain

$$U(t) \to N_0.$$

Therefore,

$$X(t) \to X_0 = (N_0, 0, 0, 0, 0) \quad \text{as } t \to \infty.$$

By LaSalle's invariance principle [22, 34], the risk-free equilibrium is globally asymptotically stable in Ω. □

Economically, this result means that if $R_0 < 1$, financial distress cannot persist in the banking system. Even if some banks are initially exposed or distressed, the system returns to the risk-free state in the long run.

5.4 Global stability of the persistent-risk equilibrium

Assume that $R_0 > 1$, so that the persistent-risk equilibrium $X^* = (U^*, E^*, D^*, R^*, L^*)$ exists in Ω with $E^* > 0$ and $D^* > 0$. Since $N(t) = N_0$, we substitute $U = N_0 - E - D - R - L$ and work on the reduced system

$$\begin{cases} \dot{E} = \beta(N_0 - E - D - R - L)D - (\mu + \gamma)E, \\ \dot{D} = \gamma E - (\delta_1 + \delta_2 + \delta_3)D, \\ \dot{R} = \delta_1 D - \theta R, \\ \dot{L} = \delta_2 D - \eta L, \end{cases} \tag{40}$$

on $\Omega_4 = \{(E, D, R, L) \in \mathbb{R}^4_+ : E + D + R + L \leq N_0\}$.

Theorem 4. If $R_0 > 1$, then the persistent-risk equilibrium X^* is globally asymptotically stable in Ω.

Proof. Define the Lyapunov function [31]

$$\mathcal{V}(Y) = \left(E - E^* - E^* \ln \tfrac{E}{E^*}\right) + \frac{\mu + \gamma}{\gamma}\left(D - D^* - D^* \ln \tfrac{D}{D^*}\right). \tag{41}$$

Since $g(u) = u - 1 - \ln u \geq 0$ for all $u > 0$, we have $\mathcal{V} \geq 0$ with $\mathcal{V}(Y^*) = 0$. Differentiating along (40) and using the equilibrium identities $\beta U^* D^* = (\mu + \gamma)E^*$ and $\gamma E^* = (\delta_1 + \delta_2 + \delta_3)D^*$, we obtain after simplification

$$\dot{\mathcal{V}} = (\mu + \gamma)E^* \left[2 + ab - \frac{ab}{c} - \frac{c}{b} - b\right], \tag{42}$$

where $a = U/U^*$, $b = D/D^*$, $c = E/E^*$. By the AM–GM inequality,

$$\frac{ab}{c} + \frac{c}{b} \geq 2\sqrt{a} \quad \text{and} \quad b + \frac{1}{b} \geq 2,$$

and applying the standard inequality $g(u) \geq 0$ to each term in (42), one obtains [31]

$$2 + ab - \frac{ab}{c} - \frac{c}{b} - b \leq 0 \quad \text{for all } a, b, c > 0, \tag{43}$$

with equality if and only if $a = b = c = 1$, i.e. $U = U^*$, $D = D^*$, $E = E^*$. Hence $\dot{\mathcal{V}} \leq 0$ on Ω_4.

On the set $\{\dot{\mathcal{V}} = 0\}$, we have $D = D^*$. Substituting into $\dot{R} = \delta_1 D - \theta R$ gives $R(t) \to R^*$ exponentially, and into $\dot{L} = \delta_2 D - \eta L$ gives $L(t) \to L^*$ exponentially. By LaSalle's invariance principle [22], every trajectory in Ω converges to X^*. □

When $R_0 > 1$, the system moves toward the persistent-risk equilibrium for any initial condition. Increasing δ_3 above δ_3^* reduces R_0 below one and brings the system back to the risk-free state.

6 Numerical simulations

Numerical simulations are presented to support the analytical results obtained in the previous sections [21, 14, 16]. The objective is to illustrate the role of the risk reproduction number R_0 and to show the effect of the two financial mechanisms introduced in the model, namely the direct intervention parameter δ_3 and the reintegration parameter η.

The simulations are performed for system (1). Unless otherwise stated, the initial condition is

$$(U(0), E(0), D(0), R(0), L(0)) = (90, 5, 3, 1, 1),$$

with $N_0 = 100$. The parameter configurations used in the numerical tests are summarized in Table 3. The first scenario corresponds to $R_0 > 1$, while the second one corresponds to $R_0 < 1$. The last two scenarios are used to examine the effects of δ_3 and η.

Table 3: Numerical scenarios used in the simulations.

Scenario	Main change	Parameter values	R_0	Figure
Persistent-risk regime	High contagion	$\beta = 1.2/N_0$	10	Fig. 2
Risk-free regime	Low contagion	$\beta = 0.01/N_0$	0.083	Fig. 3
Direct intervention	Varying δ_3	$\delta_3 \in \{0.03, \ldots, 0.15\}$	decreases with δ_3	Fig. 4
Reintegration	Varying η	$\eta > 0$	unchanged	Fig. 5

For the illustrative simulations, the baseline parameter values are

$$\gamma = 0.5, \quad \mu = 0.1, \quad \delta_1 = 0.05, \quad \delta_2 = 0.02,$$

$$\delta_3 = 0.03, \quad \theta = 0.1, \quad \eta = 0.05.$$

These illustrative parameter values are used unless otherwise specified. They were chosen to show the cases $R_0 > 1$ and $R_0 < 1$ and to study the effects of δ_3 and η. These values do not come from a specific banking data set. In future work, the model could be calibrated using data on bank capital, liquidity, nonperforming loans, and interbank exposures. These indicators

could be used to place each bank in one of the five compartments. The transition parameters could then be estimated from changes in the financial state of banks over time. The parameter β could be estimated by comparing the observed and simulated numbers of exposed and distressed banks. The calibrated model could then be tested over another period [11, 12].

A. Persistent-risk regime

We first consider the case

$$\beta = \frac{1.2}{N_0} = 0.012.$$

Then

$$\mu + \gamma = 0.6, \qquad \delta_1 + \delta_2 + \delta_3 = 0.10.$$

Therefore,

$$R_0 = \frac{\beta N_0 \gamma}{(\mu+\gamma)(\delta_1+\delta_2+\delta_3)} = \frac{0.012 \times 100 \times 0.5}{0.6 \times 0.10} = 10.$$

Since $R_0 > 1$, the system is in the persistent-risk regime. The corresponding equilibrium is

$$X^* \approx (9.99,\ 8.57,\ 42.89,\ 21.47,\ 17.07).$$

Figure 2 shows the time evolution of the five compartments. After a short transition phase, the solution approaches the persistent-risk equilibrium. The distressed class stabilizes near

$$D^* \approx 42.89.$$

This confirms that, when $R_0 > 1$, financial distress remains present in the banking network.

The largest real part of the eigenvalues is approximately

$$\Re(\lambda_{\max}) \approx -0.059.$$

This value is negative, which supports the local stability result obtained from the Routh–Hurwitz criterion. The convergence is slow because this eigenvalue is close to zero.

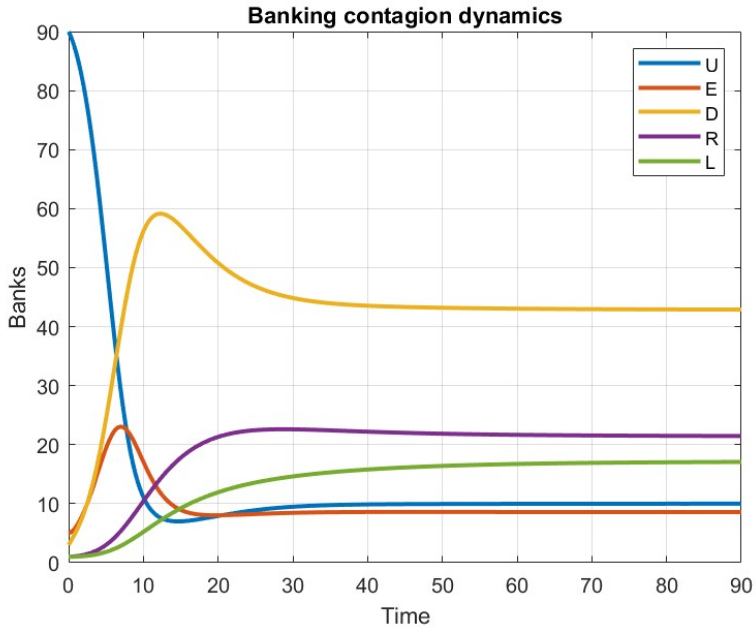


Figure 2: Time evolution of the banking system in the persistent-risk regime ($R_0 > 1$).

B. Risk-free regime

We now reduce the contagion intensity and take

$$\beta = \frac{0.01}{N_0} = 0.0001.$$

Keeping all other parameters unchanged, we obtain

$$R_0 = \frac{0.0001 \times 100 \times 0.5}{0.6 \times 0.10} \approx 0.083 < 1.$$

Thus, the system is in the risk-free regime. According to the global stability result proved in Section 5, the solution converges to the risk-free equilibrium.

Figure 3 confirms this result. The exposed and distressed classes decrease toward zero, while the unexposed class tends to N_0. Numerically, the solution approaches

$$(99.94,\ 0,\ 0.002,\ 0.007,\ 0.051),$$

which is close to the risk-free equilibrium.

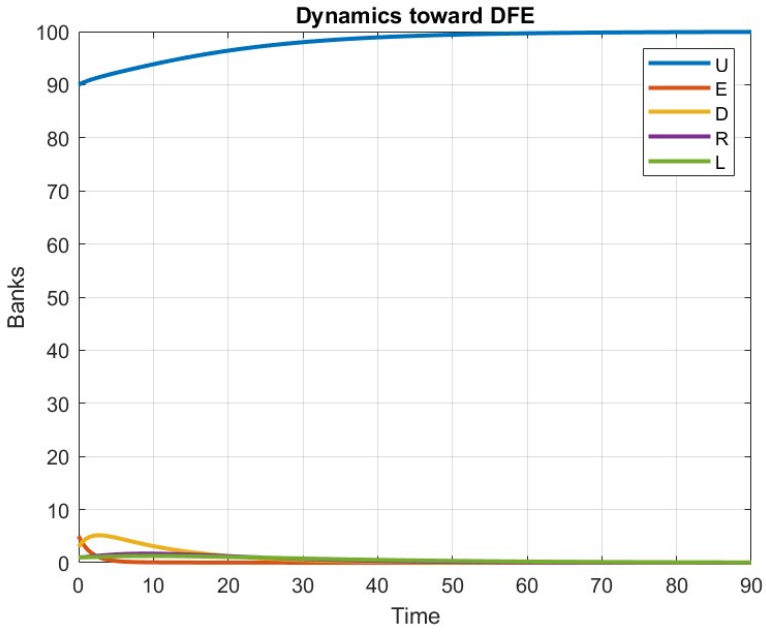


Figure 3: Time evolution of the banking system in the risk-free regime ($R_0 < 1$).

In this regime, the crisis remains limited. The maximum value of the distressed class is approximately

$$D_{\max} \approx 5.16,$$

after which $D(t)$ decreases toward zero. This agrees with the Lyapunov–LaSalle result established for $R_0 < 1$.

C. Effect of direct intervention

The parameter δ_3 represents the direct intervention mechanism from the distressed class to the healthy class. It appears in the denominator of the risk reproduction number:

$$R_0 = \frac{\beta N_0 \gamma}{(\mu + \gamma)(\delta_1 + \delta_2 + \delta_3)}.$$

Hence, increasing δ_3 decreases R_0 and reduces the propagation of financial distress.

The critical value of δ_3 is obtained from $R_0 = 1$ as follows:

$$\delta_3^* = \frac{\beta N_0 \gamma}{\mu + \gamma} - (\delta_1 + \delta_2). \tag{44}$$

For the persistent-risk parameter set, we get

$$\delta_3^* = \frac{0.012 \times 100 \times 0.5}{0.6} - (0.05 + 0.02) = 0.93.$$

Thus, values of δ_3 below 0.93 keep the system in the persistent-risk regime, while values above this level move the system toward the risk-free regime.

Figure 4 shows the effect of δ_3 on the distressed compartment. When δ_3 increases, the level of distress decreases. In particular, the long-run value of $D(t)$ decreases from about 43 for $\delta_3 = 0.03$ to nearly 30 for $\delta_3 = 0.15$. This confirms the stabilizing role of direct intervention.

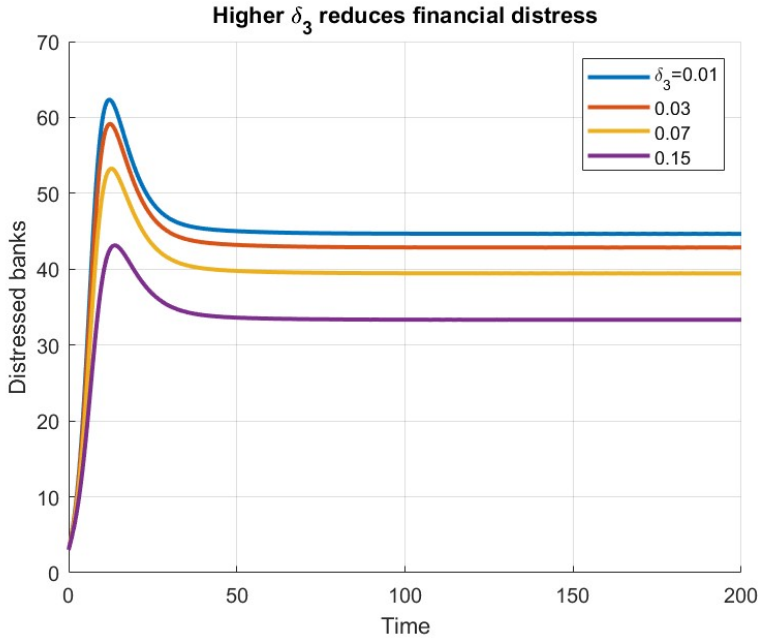


Figure 4: Effect of the direct intervention parameter δ_3 on the distressed compartment.

From a financial point of view, increasing δ_3 means that regulators intervene before the distress spreads further in the network. This mechanism reduces systemic risk directly, because it acts on R_0.

D. Effect of reintegration from liquidation

The parameter η represents the reintegration of liquidated institutions into the unexposed class. Unlike δ_3, it does not appear in the expression of R_0. Therefore, η does not directly determine whether the system is in the risk-free or persistent-risk regime. However, it affects the final distribution of banks.

From the equilibrium equations, we have

$$L^* = \frac{\delta_2}{\eta} D^*.$$

This relation shows that increasing η reduces the long-term number of liquidated institutions.

Figure 5 shows the effect of η on the liquidated class. When η increases, the liquidated compartment decreases faster and reaches a lower final level. For example, when $\eta = 0.05$ and $D^* \approx 42.89$, we obtain

$$L^* \approx \frac{0.02}{0.05} \times 42.89 \approx 17.16,$$

which is consistent with the numerical equilibrium.

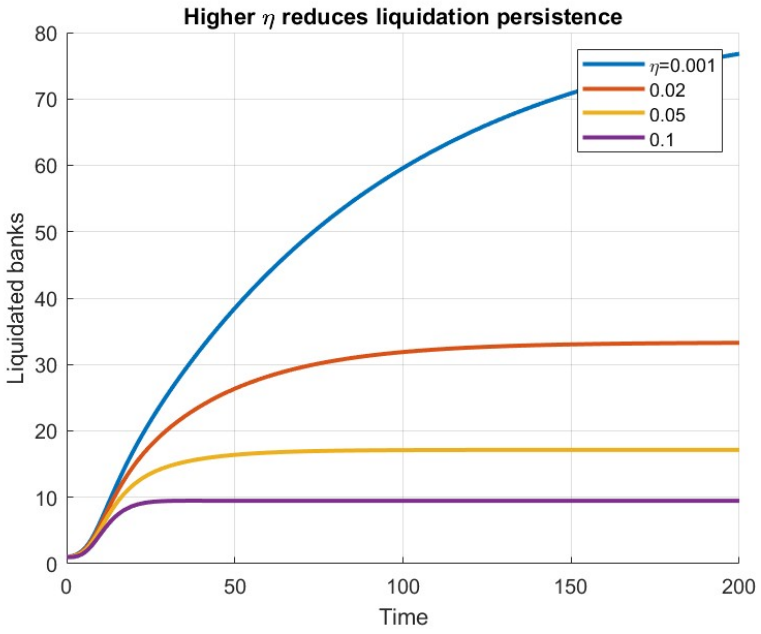


Figure 5: Effect of the reintegration parameter η on the liquidated compartment.

This result shows that η mainly improves the reintegration process after liquidation. It reduces the number of liquidated banks and accelerates reintegration, but it does not reduce R_0 directly.

E. Discussion

The numerical results confirm the analytical role of R_0. When $R_0 < 1$, the exposed and distressed compartments vanish and the system approaches the risk-free equilibrium. When $R_0 > 1$, the system approaches a persistent-risk state in which distress remains present.

The parameter δ_3 acts directly on R_0, whereas η improves reintegration without changing R_0. In practice, a higher value of δ_3 may represent faster and more effective early supervisory actions, such as the activation of recovery plans, recapitalization, restrictions on distributions, reductions in risky exposures, or changes in management [3]. These actions could be triggered by deteriorating capital, liquidity, asset-quality, or interbank-exposure indicators. The parameter η may instead represent the transfer of viable activities to an acquiring institution or a bridge bank [32]. Thus, the model represents the combined speed and effectiveness of regulatory intervention rather than prescribing a single policy instrument. These simulations confirm the analytical results but do not constitute an empirical validation against an observed banking crisis. The model assumes that all banks have the same size and interact in the same way. However, large banks and highly connected banks may have a stronger effect on the spread of financial risk. This is related to the "Too Big to Fail" problem, since the distress of a systemically important bank may affect a large part of the banking system. The present model describes the average behavior of banks and does not distinguish systemically important institutions. Bank size and interbank connections may be included in future work [1].

7 Local and global sensitivity analysis

In this section, we study the sensitivity of the risk reproduction number R_0 with respect to the model parameters. This analysis helps identify the parameters that have the strongest influence on the propagation of systemic risk in the banking network.

Following [6], the normalized sensitivity index of R_0 with respect to a parameter p is defined by

$$S_p = \frac{\partial R_0}{\partial p}\frac{p}{R_0}. \tag{45}$$

Recall that

$$R_0 = \frac{\beta N_0 \gamma}{(\mu+\gamma)(\delta_1+\delta_2+\delta_3)}. \tag{46}$$

By direct differentiation, we obtain

$$\begin{aligned} S_\beta &= 1, & S_{N_0} &= 1, \\ S_\gamma &= \frac{\mu}{\mu+\gamma}, & S_\mu &= -\frac{\mu}{\mu+\gamma}, \\ S_{\delta_j} &= -\frac{\delta_j}{\delta_1+\delta_2+\delta_3}, & j &= 1,2,3. \end{aligned} \tag{47}$$

The local sensitivity analysis considers one parameter at a time. To obtain a more complete comparison, we also performed a global sensitivity analysis using Latin hypercube sampling (LHS) and partial rank correlation coefficients (PRCCs), following Marino et al. [23]. The seven parameters appearing in R_0 were varied simultaneously and independently within $\pm 20\%$ of their reference values. A total of 5000 parameter sets were generated. The results are summarized in Table 4.

Table 4: Local and global sensitivity results for the risk reproduction number R_0.

Parameter	Reference value	Local index	PRCC	p-value
β	0.012	1.0000	0.9653	< 0.001
N_0	100	1.0000	0.9651	< 0.001
γ	0.5	0.1667	0.5165	< 0.001
μ	0.1	-0.1667	-0.4976	< 0.001
δ_1	0.05	-0.5000	-0.8710	< 0.001
δ_2	0.02	-0.2000	-0.5751	< 0.001
δ_3	0.03	-0.3000	-0.7178	< 0.001

Figure 6 presents the PRCC values obtained from the global sensitivity analysis.

The local and global analyses give the same directions of influence. The parameters β, N_0, and γ have positive effects on R_0, whereas μ, δ_1, δ_2, and δ_3 have negative effects. All PRCC values are statistically significant at $p < 0.001$.

The parameters β and N_0 have the strongest positive influence, with PRCC values equal to 0.9653 and 0.9651, respectively. This shows that an increase in contagion intensity or in the size

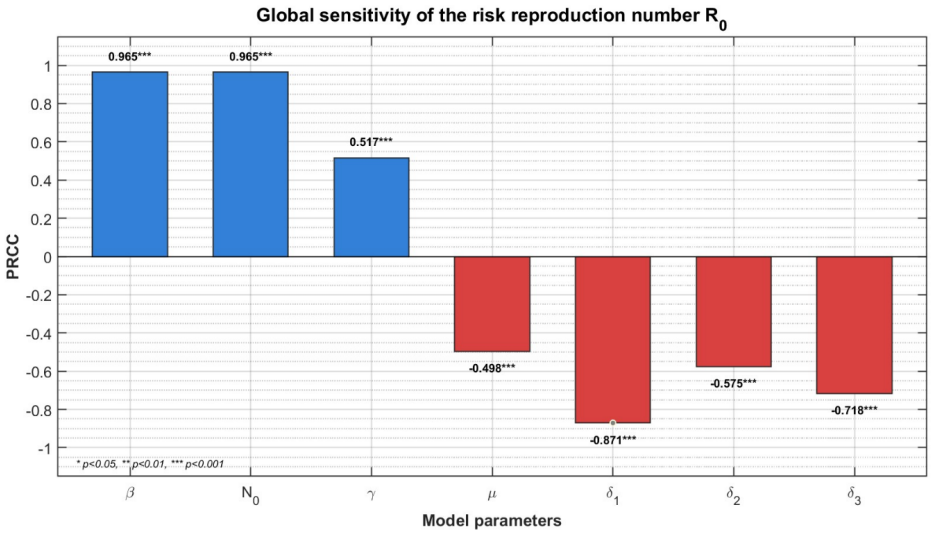


Figure 6: PRCCs of the risk reproduction number R_0 obtained from 5000 Latin hypercube samples. The parameters were varied independently within $\pm 20\%$ of their reference values. All PRCCs are statistically significant ($p < 0.001$).

of the banking system strongly increases the propagation of systemic risk. The parameter γ also increases R_0, but its influence is smaller.

Among the parameters with negative effects, δ_1 has the largest absolute PRCC value, followed by δ_3 and δ_2. Thus, these intervention mechanisms play an important role in reducing financial contagion. In particular, the negative PRCC of δ_3 confirms that direct early intervention can reduce the propagation of financial distress. The parameter μ also reduces R_0, but it has the smallest global influence among the parameters considered.

Since the parameter distributions have not yet been estimated from real banking data, the intervals used in this analysis are exploratory. Future work will consider parameter ranges estimated from banking data and the global sensitivity of other model outputs.

8 Conclusion

Several mathematical models of financial contagion are inspired by epidemiological models, with some adjustments to represent the functioning of banking systems. In this work, we followed a different approach. We kept the compartmental structure but modified two points that we consider important. First, a distressed bank does not always follow the complete recovery process. Regulators can intervene early and return it to a sound condition. Second, liquidation is not

necessarily the final step, since viable banking activities may return to the system after acquisition or reorganization. These observations led us to introduce the parameters δ_3 and η into the UEDRL model.

The mathematical analysis allowed us to determine the role of the risk reproduction number in separating the two possible regimes. We proved that the risk-free equilibrium is globally asymptotically stable when $R_0 < 1$. When $R_0 > 1$, a persistent-risk equilibrium exists, and we established its local and global asymptotic stability. The numerical simulations confirm these analytical results. The local and global sensitivity analyses also show that the contagion rate and the size of the banking system have the strongest positive influence on R_0, while the intervention parameters δ_1, δ_2, and δ_3 have negative effects. In particular, δ_3 acts directly on R_0, whereas η changes the long-run distribution of banks without affecting this number.

From a regulatory perspective, δ_3 may represent the combined speed and effectiveness of early supervisory actions, such as the activation of recovery plans, recapitalization, restrictions on distributions, reductions in risky exposures, or changes in management. However, the model remains a simplified representation of a banking system. It assumes a closed and homogeneous population, mass-action contagion, constant parameters, and illustrative parameter values. It does not explicitly account for differences in bank size, heterogeneous interbank connections, or time-dependent regulatory decisions. Future work will therefore consider calibration with bank-level data, heterogeneous banking networks, time-dependent interventions, and empirical validation over different periods.

A Algebraic verification of the Routh–Hurwitz conditions

We verify here the two Routh–Hurwitz conditions used in Section 5.2. Let

$$s = \alpha + q + \Delta, \qquad r = \theta + \eta, \qquad z = \theta\eta,$$

and define

$$b = q(\Delta + \gamma), \qquad c = \gamma q(\delta_1 + \delta_2), \qquad d = \gamma q(\delta_1\eta + \delta_2\theta).$$

The coefficients of the characteristic polynomial can then be written as

$$a_1 = s + r, \qquad a_2 = b + sr + z, \qquad a_3 = br + sz + c, \qquad a_4 = bz + d. \tag{48}$$

Indeed, the characteristic polynomial of the reduced Jacobian admits the following compact form:

$$p(\lambda) = (\lambda+\theta)(\lambda+\eta)\left[\lambda^2 + s\lambda + b\right] + \gamma q\left[\delta_1(\lambda+\eta) + \delta_2(\lambda+\theta)\right]. \tag{49}$$

Expanding (49) gives precisely the four coefficients in (48).

For the first Routh–Hurwitz determinant, we obtain

$$\begin{aligned} a_1a_2 - a_3 &= (s+r)(b+sr+z) - (br+sz+c) \\ &= sb + s^2r + sr^2 + rz - c. \end{aligned} \tag{50}$$

Using the definitions of b and c, this becomes

$$a_1a_2 - a_3 = s^2r + sr^2 + rz + sq\Delta + \gamma q\left(s - \delta_1 - \delta_2\right). \tag{51}$$

Moreover,

$$s - \delta_1 - \delta_2 = \alpha + q + \delta_3 > 0.$$

Hence every term on the right-hand side of (51) is strictly positive, and therefore

$$a_1a_2 - a_3 > 0.$$

For the second determinant, let

$$H_1 = a_1a_2 - a_3.$$

Then

$$a_1a_2a_3 - a_3^2 - a_1^2a_4 = a_3H_1 - a_1^2a_4. \tag{52}$$

Substituting (48) and (50) into (52) gives

$$\begin{aligned} H_2 &:= a_1a_2a_3 - a_3^2 - a_1^2a_4 \\ &= (br+sz+c)\left(sb + s^2r + sr^2 + rz - c\right) \\ &\quad - (s+r)^2(bz+d). \end{aligned} \tag{53}$$

To verify the sign of H_2, we substitute

$$s = \alpha + \Delta + q, \qquad \Delta = \delta_1 + \delta_2 + \delta_3,$$

and collect the resulting expression according to the powers of q. This gives the exact decomposition

$$H_2 = P_0 + qP_1 + q^2P_2 + q^3P_3, \tag{54}$$

where

$$P_0 = \theta\eta(\theta+\eta)(\alpha+\Delta)(\alpha+\Delta+\theta)(\alpha+\Delta+\eta) > 0. \tag{55}$$

The coefficients P_1, P_2, P_3 are obtained by exact polynomial collection after using

$$\alpha = \mu + \gamma, \qquad \Delta = \delta_1 + \delta_2 + \delta_3.$$

Their signs were checked without assigning numerical values to the parameters. For completeness, the exact symbolic verification is given below in SageMath syntax:

```
S.<q> = PolynomialRing(
    PolynomialRing(QQ,
    names=('mu','gamma','d1','d2','d3','theta','eta'))
)
mu,gamma,d1,d2,d3,theta,eta = S.base_ring().gens()

alpha = mu + gamma
Delta = d1 + d2 + d3
s = alpha + Delta + q
r = theta + eta
z = theta*eta

b = q*(Delta + gamma)
c = gamma*q*(d1 + d2)
d = gamma*q*(d1*eta + d2*theta)

a1 = s + r
a2 = b + s*r + z
a3 = b*r + s*z + c
a4 = b*z + d

H2 = a1*a2*a3 - a3^2 - a1^2*a4

for k in range(4):
    Pk = H2[k]
    assert all(v > 0 for v in Pk.dict().values())
```

After collection, P_1, P_2, P_3 contain, respectively, 169, 132, and 41 distinct monomials, and all their coefficients are strictly positive. Hence,

$$P_i > 0, \qquad i = 0, 1, 2, 3.$$

Since $q = \beta D^* > 0$, relation (54) yields

$$a_1 a_2 a_3 - a_3^2 - a_1^2 a_4 > 0.$$

Therefore, the two remaining Routh–Hurwitz conditions are satisfied.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] Acemoglu, D., Ozdaglar, A. and Tahbaz-Salehi, A., *Systemic risk and stability in financial networks*, Amer. Econ. Rev. 105(2) (2015), 564–608.

[2] Ait Tamerz, H., Allali, K. and Meskaf, A., *Contagion in the banking ecosystem: fractional order modelling*, Commun. Math. Biol. Neurosci. 2024 (2024), 123.

[3] Angeloni, I., *Crisis management in the banking union: Overview and early experience*, European Central Bank, ECB Banking Supervision, 29 June 2017. Available at: https://www.bankingsupervision.europa.eu/press/speeches/date/2017/html/ssm.sp170629.en.html.

[4] Brunnermeier, M. K., *Deciphering the liquidity and credit crunch 2007–2008*, J. Econ. Perspect. 23(1) (2009), 77–100.

[5] Brunnermeier, M. K. and Sannikov, Y., *A macroeconomic model with a financial sector*, Amer. Econ. Rev. 104(2) (2014), 379–421.

[6] Chitnis, N., Hyman, J. M. and Cushing, J. M., *Determining important parameters in the spread of malaria through the sensitivity analysis of a mathematical model*, Bull. Math. Biol. 70(5) (2008), 1272–1296.

[7] Coddington, E. A. and Levinson, N., *Theory of ordinary differential equations*, McGraw-Hill, New York, 1955.

[8] Cont, R. and Minca, A., *Recovering portfolio default intensities implied by CDO quotes*, Math. Finance 23(1) (2013), 94–121.

[9] Diekmann, O., Heesterbeek, J. A. P. and Metz, J. A. J., *On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations*, J. Math. Biol. 28 (1990), 365–382.

[10] Ding, Z., Yan, H., Chen, Y. and Jiang, C., *Risk contagion in interbank lending networks: a multi-agent-based modeling and simulation perspective*, Expert Syst. Appl. 255 (2024), 124847.

[11] European Banking Authority, *Pillar 3 Data Hub*, European Data Access Portal, 2026. Available at: https://www.eba.europa.eu/risk-and-data-analysis/pillar-3-data-hub (accessed 10 August 2026).

[12] European Central Bank, *Supervisory data*, ECB Banking Supervision, 2026. Available at: https://www.bankingsupervision.europa.eu/framework/statistics/html/index.en.html (accessed 10 August 2026).

[13] Farah, E. M., Amine, S. and Allali, K., *Dynamics of a time-delayed two-strain epidemic model with general incidence rates*, Chaos Solitons Fractals 153 (2021), 111527.

[14] Farah, E. M., Amine, S., Hajhouji, Y. and Allali, K., *Global dynamics of a time-delayed two-strain epidemic model with general incidence rates and vaccination*, Model. Earth Syst. Environ. 10 (2024), 6407–6428.

[15] Hałaj, G. and Hipp, R., *Decomposing systemic risk: The roles of contagion and common exposures*, J. Financ. Stab. 80 (2025), 101451.

[16] Irakoze, I., Nahayo, F., Ikpe, D., Gyamerah, S. A. and Viens, F., *Mathematical modeling and stability analysis of systemic risk in the banking ecosystem*, J. Appl. Math. 2023 (2023), e5628621.

[17] Irakoze, I., Ikpe, D., Nahayo, F. and Gyamerah, S. A., *Time dynamics of systemic risk in banking networks: a UEDR-PDE approach*, AppliedMath 5(2) (2025), 54.

[18] Jiang, E. X., Matvos, G., Piskorski, T. and Seru, A., *Monetary tightening and U.S. bank fragility in 2023: mark-to-market losses and uninsured depositor runs?*, J. Financ. Econ. 159 (2024), 103899.

[19] Jobst, A., *Quantifying systemic risk*, Oxf. Rev. Econ. Policy 29(3) (2013), 512–542.

[20] Khalil, H. K., *Nonlinear systems*, 3rd edn., Prentice Hall, Upper Saddle River, 2002.

[21] Khyar, O., Meskaf, A. and Allali, K., *Mathematic analysis of a SIHV COVID-19 pandemic model taking into account a vaccination strategy*, in: R. P. Mondaini (ed.), *Trends in Biomathematics*, Springer, Cham, (2022), 193–223.

[22] LaSalle, J. P., *Some extensions of Liapunov's second method*, IRE Trans. Circuit Theory 7(4) (1960), 520–527.

[23] Marino, S., Hogue, I. B., Ray, C. J. and Kirschner, D. E., *A methodology for performing global uncertainty and sensitivity analysis in systems biology*, J. Theor. Biol. 254 (2008), 178–196.

[24] Meskaf, A., Khyar, O., Danane, J. and Allali, K., *Global stability analysis of a two-strain epidemic model with non-monotone incidence rates*, Chaos Solitons Fractals 133 (2020), 109647.

[25] Mishkin, F. S., *Over the cliff: from the subprime to the global financial crisis*, J. Econ. Perspect. 25(1) (2011), 49–70.

[26] Mourad, N., Taamouti, A. and Hamdoune, M., *Modeling and mathematical analysis of liquidity risk contagion in the banking system*, J. Appl. Math. 2022 (2022), 5382153.

[27] Nagumo, M., *Über die Lage der Integralkurven gewöhnlicher Differentialgleichungen*, Proc. Phys. Math. Soc. Jpn. 24 (1942), 551–559.

[28] Narayan, S. and Kumar, D., *Macroprudential policy and systemic risk in G20 nations*, J. Financ. Stab. 75 (2024), 101340.

[29] Nowaczyk, N. and O'Halloran, S., *Computing the impact of central clearing on systemic risk*, Front. Artif. Intell. 7 (2024), 1138611.

[30] Petrone, D. and Latora, V., *A dynamic approach merging network theory and credit risk techniques to assess systemic risk in financial networks*, Sci. Rep. 8 (2018), 5561.

[31] Shuai, Z. and van den Driessche, P., *Global stability of infectious disease models using Lyapunov functions*, SIAM J. Appl. Math. 73(4) (2013), 1513–1532.

[32] Single Resolution Board, *Resolution tools*, European Union. Available at: https://www.srb.europa.eu/en/content/tasks-tools (accessed 10 August 2026).

[33] Smith, H. L., *Monotone dynamical systems*, American Mathematical Society, Providence, 1995.

[34] Thieme, H. R., *Mathematics in population biology*, Princeton University Press, Princeton, 2003.

[35] van den Driessche, P. and Watmough, J., *Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission*, Math. Biosci. 180 (2002), 29–48.

Iranian Journal of Numerical Analysis and Optimization
Vol. 17, No. 1, 2027, pp 529–569
https://doi.org/10.22067/ijnao.2026.100313.1935
https://ijnao.um.ac.ir/

CC BY NC

Research Article

Mathematical modeling, global stability, sensitivity analysis, and optimal control of merchant adoption dynamics in e-commerce

H. Hassouni, A. El Bhih*, O. Aarabate and O. Balatif

Abstract

In this paper, we introduce a novel mathematical model to analyze the dynamics of e-commerce adoption among merchants. The model captures the key factors influencing adoption behaviors and examines both the spread of e-commerce and resistance to its uptake.

Using the next-generation matrix method, we establish the model's well-posedness and derive the basic reproduction number R_0. Two equilibrium points are identified: The rejection

*Corresponding author

Received 6 August 2026; revised 11 September 2026; accepted 15 September 2026

Habib Hassouni
Laboratory of Fundamental Mathematics and their Applications, Department of Mathematics, Faculty of Sciences El Jadida, Chouaib Doukkali University, El Jadida, Morocco. e-mail: habibhassouni85@gmail.com

Amine El Bhih
Multidisciplinary Research and Innovation Laboratory (LPRI), Moroccan School of Engineering Sciences, Casablanca, Morocco. e-mail: elbhihamine@gmail.com

Oumaima Aarabate
Laboratory of Fundamental Mathematics and their Applications, Department of Mathematics, Faculty of Sciences El Jadida, Chouaib Doukkali University, El Jadida, Morocco. e-mail: oumaimaaarabat@gmail.com

Omar Balatif
Laboratory of Fundamental Mathematics and their Applications, Department of Mathematics, Faculty of Sciences El Jadida, Chouaib Doukkali University, El Jadida, Morocco. e-mail: balatif.maths@gmail.com

How to cite this article

Hassouni, H., El Bhih, A., Aarabate, O. and Balatif, O., Mathematical modeling, global stability, sensitivity analysis, and optimal control of merchant adoption dynamics in e-commerce. *Iran. J. Numer. Anal. Optim.*, 2027; 17(1): 529-569. https://doi.org/10.22067/ijnao.2026.100313.1935

equilibrium E_0, representing a state of nonadoption, and the widespread adoption equilibrium E_*, representing a state of large-scale e-commerce adoption. A global stability analysis, based on LaSalle's invariance principle and Lyapunov functions, demonstrates that E_0 is globally asymptotically stable if $R_0 \leq 1$, while E_* is globally asymptotically stable when $R_0 > 1$.

Furthermore, the model explores the drivers of e-commerce diffusion and factors contributing to resistance. We formulate the model as an optimal control problem to address these dynamics, introducing control variables aimed at minimizing resistance and promoting widespread adoption. Cost-effectiveness analysis is also performed to evaluate the efficiency of each intervention strategy.

Finally, numerical simulations provide insights into the effectiveness of various policy interventions, illustrating the strategies required to achieve optimal levels of e-commerce adoption.

AMS subject classifications (2020): Primary 34D20; Secondary 49J15, 49K15, 92D30.

Keywords: Mathematical modeling; Stability analysis; E-commerce; Pontryagin maximum principle; Optimal control; Sensitivity analysis.

A Introduction

E-commerce is widely recognized as a pivotal driver of economic growth at both national and international levels. It provides a dynamic platform for businesses to showcase and sell their products to both domestic and global markets, while offering consumers the convenience of purchasing goods and services from virtually any location at any time. As a result, the growth of e-commerce is expected to significantly reduce the time, energy, and effort required to complete transactions.

Beyond economic benefits, the expansion of e-commerce also yields broader societal impacts. For instance, the shift to digital transactions contributes to lower pollution levels by reducing the need for physical commuting, helps limit the spread of infectious diseases by minimizing in-person interactions, and decreases the frequency of traffic-related accidents by promoting remote shopping alternatives.

E-commerce refers to the buying, selling, and provision of goods and services over the Internet. This process is supported by a diverse range of digital technologies, including computers, smartphones, tablets, and, increasingly, wearable devices. These technologies collectively form a robust infrastructure that enables the seamless execution of commercial activities through online platforms, ensuring real-time communication, secure transactions, and rapid delivery services.

In recent years, the global proliferation of e-commerce has been nothing short of remarkable. However, effective participation in this digital economy requires both reliable communication infrastructure and a high level of proficiency in managing commercial websites. Depending on the nature of interaction, various e-commerce models have emerged, such as business-to-consumer (B2C), business-to-business (B2B), consumer-to-consumer (C2C), and others [6]. Each of these models addresses different segments of the market and plays a complementary role in the digital ecosystem.

The historical development of e-commerce has been shaped by several key milestones. One of the earliest milestones was the introduction of electronic data interchange in the 1960s, which allowed organizations to exchange business documents electronically. In 1969, CompuServe was founded, offering innovative services such as email and online chat that laid the groundwork for future online communication.

A major breakthrough came in 1979, when Michael Aldrich pioneered the concept of online purchasing using videotex technology, marking one of the first true implementations of electronic commerce. The 1990s then witnessed an exponential acceleration in the field: The launch of the Netscape Navigator browser in 1994 significantly enhanced web accessibility for the general public, while the founding of Amazon.com and eBay in 1995 revolutionized the way goods were bought and sold online.

Further important developments include Dell's implementation of direct online sales in 1996 and Alibaba's entry into the market in 1999, which positioned China as a major player in the global e-commerce landscape. More recently, in 2011, Facebook introduced targeted advertising based on user behavior, effectively integrating social media and e-commerce and giving rise to personalized digital marketing strategies [24].

These technological advancements, combined with changing consumer behavior and increased digital literacy, continue to shape the future trajectory of e-commerce as a dominant mode of global trade and interaction.

Related work

The modeling of infectious diseases provides a comprehensive and rigorous framework for analyzing pathogen transmission mechanisms, forecasting the future course of an epidemic, and evaluating strategies for epidemic control. Interestingly, e-commerce models share parallels with biological epidemiological models, as they elucidate transmission processes such as word-of-mouth

or advertising campaigns and the temporal evolution of states within a consumer population, including awareness, interest, and adoption.

Applying these principles to e-commerce allows for greater precision in simulating product diffusion, innovation adoption, and marketing campaign effects. For instance, modeling the spread of a new technological product can help predict its market penetration over time. This analytical approach not only enhances our understanding of the fundamental mechanisms driving market volatility but also enables more accurate prediction of consumer behavior in response to various stimuli.

Integrating epidemiological models into the mathematical and economic modeling of e-commerce presents an opportunity to optimize commercial strategies and anticipate trends in the rapidly evolving digital marketplace.

In [25], the authors modeled e-commerce systems using queuing theory to evaluate the efficiency of a company's transition from a traditional business model to a digital framework. They developed a system of differential equations to analyze key variables, including order flow, processing times, and operational costs. This methodological approach facilitates a precise assessment of the transition's profitability and cost efficiency by accounting for the complexities inherent in transactional dynamics. The study's findings provide valuable insights into the economic implications of digital transformation.

In [15], a compartmental epidemic model was proposed to analyze the propagation of rumours in e-commerce, examining the stability conditions of the equilibrium points. The epidemic threshold R_0 is critical in determining whether rumours become endemic or die out, thereby influencing consumer trust in online transactions.

Li [19] presented a dynamical clustering method for electronic commerce systems, leveraging optimization and game theory to identify and form clusters efficiently. A two-stage game-theoretic framework is used, where the equilibrium contributions of leaders are computed in a sequential-move setting, leading to an efficient optimization of cluster configurations. The proposed dynamical system guarantees rapid convergence with computational complexity nearing linear time for sparse networks, making it suitable for large-scale applications.

In [1], the authors examined how e-commerce policy can enhance inclusive development in the global economy. They highlighted the importance of improving formal institutions, fostering SME participation, and supporting less-developed countries in e-commerce.

Lv and Chen [21] established a time-delayed panic-spreading model with stability analysis and optimal control. They studied the local and global stability of equilibria and used optimal control to minimize the spread of panic through numerical simulations.

Those interested in exploring additional references in this field can refer to [11, 9, 10].

Problem statement

E-commerce has fundamentally transformed global trade; however, the dynamic behavior of merchants toward its adoption remains insufficiently explored in the existing literature. In particular, previous studies have not adequately addressed how merchants who refuse to use e-commerce may influence those who choose to adopt it. To address this gap, we introduce a mathematical model to analyze the dynamic behavior of merchants in relation to e-commerce adoption and rejection. The proposed model considers a compartmental framework comprising five categories: Merchants who are not currently engaged in e-commerce but may potentially adopt it; merchants who actively refuse to adopt e-commerce platforms; merchants who currently use e-commerce in their business activities; merchants who have temporarily discontinued their use of e-commerce; and merchants who have permanently ceased using e-commerce.

The model describes the transitions and interactions among these compartments, thereby providing insights into the factors influencing e-commerce adoption and rejection among merchants. We establish the well-posedness of the model and derive the basic reproduction number R_0 using the next-generation matrix method. Two equilibrium points are identified: The rejection-free equilibrium E_0, representing a state in which e-commerce rejection is absent, and the positive equilibrium E_*, representing the long-term coexistence of the different merchant behavioral states, including e-commerce adoption and rejection.

A global stability analysis is then conducted using Lyapunov functions and LaSalle's invariance principle. We establish that E_0 is globally asymptotically stable when $R_0 \leq 1$, whereas E_* is globally asymptotically stable when $R_0 > 1$. To promote e-commerce adoption and reduce merchant rejection, the model is subsequently formulated as an optimal control problem. Three time-dependent control strategies are introduced to reduce the number of merchants refusing e-commerce and to promote its adoption. The Pontryagin maximum principle is employed to characterize the corresponding optimal control strategies.

Numerical simulations are performed to evaluate the effectiveness of the proposed interventions and to illustrate the optimal strategies for enhancing e-commerce adoption while reducing merchant rejection.

Furthermore, Hassouni, El Bhih, and Balatif [13] presented an application of mathematical modeling and optimal control strategies to mitigate the spread of coronavirus, illustrating the versatility and effectiveness of control theory in epidemiological applications. Additional significant contributions to the field of control theory can be found in [7, 17, 2, 4, 16, 8].

The remainder of this paper is organized as follows. Section B introduces the mathematical model developed to analyze merchant behavior in the context of e-commerce. Section C presents the stability analysis of the equilibrium points, including both local and global stabil-

ity. A sensitivity analysis of the parameters appearing in the basic reproduction number is also conducted to identify the most influential factors affecting merchant behavior and to provide insights for the subsequent optimal control analysis. In Section D, the optimal control problem is formulated, and the existence and characterization of the proposed controls are established. Section E presents numerical simulations of the model and its different compartments to evaluate the effectiveness of the proposed intervention strategies. Finally, concluding remarks and possible perspectives are presented in Section F.

B Formulation of the mathematical model

This study introduces a novel nonlinear compartmental model, referred to as the $SAWPD$ model, designed to capture the dynamics of e-commerce adoption and disengagement among merchants. The model categorizes the merchant population into five distinct compartments based on their behavioral state with respect to e-commerce. These compartments help to systematically analyze the transition of merchants from traditional commerce to e-commerce and, conversely, the potential regression or withdrawal due to various challenges.

B.1 Mathematical framework of the $SAWPD$ model

The $SAWPD$ model segments the total population of merchants into the following five compartments:

- **Compartment** S: This group includes traditional merchants who have not yet adopted e-commerce. Their reluctance typically stems from a lack of information, insufficient digital literacy, or unfamiliarity with online business models. These merchants represent the entry point into the system.
- **Compartment** A: This compartment consists of merchants who explicitly refuse to adopt e-commerce. Their refusal is often based on negative perceptions, fear of change, or skepticism about the benefits and reliability of digital platforms.
- **Compartment** W: Merchants in this group are active e-commerce users. They have successfully integrated digital tools into their business processes and use online platforms to sell their goods and services. This is the desired state that policy interventions aim to maximize.

- **Compartment** P: This group consists of merchants who have temporarily suspended their e-commerce activities. The suspension may be due to technical issues such as internet outages, platform maintenance, or unresolved user experience problems. These merchants are potentially recoverable with adequate support.

- **Compartment** D: This final compartment represents merchants who have permanently withdrawn from using e-commerce. This decision is typically the result of exposure to cybercrime, data breaches, or persistent technical and operational difficulties. These cases represent a permanent loss to the digital economy.

The transitions between these compartments are governed by a system of nonlinear differential equations, which incorporate key behavioral and technical factors influencing adoption and withdrawal. This model provides a robust analytical framework to study the effectiveness of awareness campaigns, cybersecurity initiatives, and technical support programs in fostering digital transformation among merchants.

$$\begin{cases} \frac{dS(t)}{dt} = \Lambda - \beta S(t)A(t) - \alpha S(t) - \mu S(t), \\ \frac{dA(t)}{dt} = \beta S(t)A(t) - \gamma A(t) - \mu A(t), \\ \frac{dW(t)}{dt} = \alpha S(t) + \gamma A(t) + \theta P(t) - \eta W(t) - \delta W(t) - \mu W(t), \\ \frac{dP(t)}{dt} = \eta W(t) - \theta P(t) - \mu P(t), \\ \frac{dD(t)}{dt} = \delta W(t) - \mu D(t), \end{cases} \tag{1}$$

with the initial conditions

$$S(0) \geq 0, \quad A(0) \geq 0, \quad W(0) \geq 0, \quad P(0) \geq 0, \quad D(0) \geq 0,$$

and where $N(t)$ denotes the total number of merchants at time t, given by

$$N(t) = S(t) + A(t) + W(t) + P(t) + D(t).$$

Table 1 summarizes the key parameters used in the mathematical model for e-commerce adoption. Each parameter represents a specific behavioral or structural component influencing merchants' decisions regarding e-commerce uptake or refusal.

The structure and transitions of the proposed e-commerce adoption model are illustrated in Figure 1. The diagram shows the main compartments and the possible transitions between them, which are governed by the corresponding rates of adoption, refusal, recovery, reintegration, and exit.

Table 1: Model parameters for e-commerce adoption among merchants

Parameter	Description
Λ	The number of new merchants entering the market.
α	The proportion of merchants willing to adopt e-commerce.
γ	The proportion of merchants who initially rejected e-commerce but later transitioned to adopting it.
μ	The proportion of merchants who have permanently ceased trading activities.
θ	The proportion of merchants who resumed using e-commerce after overcoming previous challenges.
β	The proportion of merchants who transitioned to rejecting e-commerce due to negative perceptions influenced by external factors.
δ	The proportion of merchants who decided to discontinue using e-commerce as a result of exposure to cybercrime.
η	The proportion of merchants who temporarily halted their use of e-commerce due to encountering specific problems.

B.2 Model's fundamental characteristics

Theorem 1. If the initial conditions satisfy

$$S(0) \geq 0, \quad A(0) \geq 0, \quad W(0) \geq 0, \quad P(0) \geq 0, \quad D(0) \geq 0,$$

then the solutions $S(t)$, $A(t)$, $W(t)$, $P(t)$, and $D(t)$ of system (1) remain nonnegative for all $t \geq 0$.

Proof. From the first equation in system (1), we have

$$\frac{dS(t)}{dt} + \big(\beta A(t) + \alpha + \mu\big)S(t) \geq \Lambda.$$

Since $\Lambda \geq 0$, it follows that

$$\frac{dS(t)}{dt} + \big(\beta A(t) + \alpha + \mu\big)S(t) \geq 0.$$

Multiplying both sides by the integrating factor

$$\exp\left(\int_0^t (\beta A(h) + \alpha + \mu)dh\right),$$

we get

$$\frac{d}{dt}\left(S(t)\exp\left(\int_0^t (\beta A(h) + \alpha + \mu)dh\right)\right) \geq 0, \quad \text{for all } t \geq 0.$$

Integrating yields

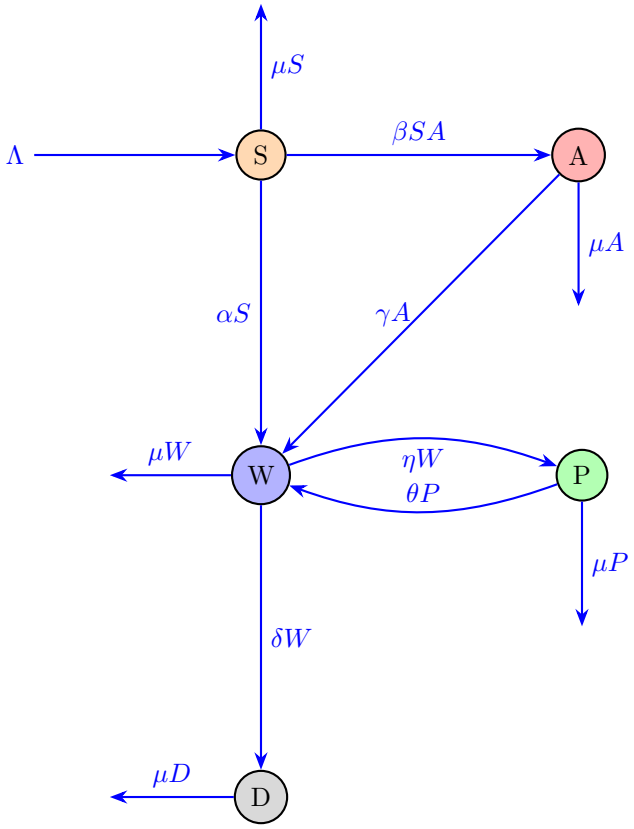


Figure 1: E-commerce merchant flow model

$$S(t) \geq S(0) \exp\left(-\int_0^t (\beta A(h) + \alpha + \mu) dh\right) \geq 0.$$

By applying similar arguments to the other equations, we conclude

$$A(t) \geq 0, \quad W(t) \geq 0, \quad P(t) \geq 0, \quad D(t) \geq 0, \quad \text{for all } t \geq 0.$$

□

B.2.1 Invariant region

In this section, we establish that all solutions of system (1) are bounded within a positively invariant region.

Theorem 2. Define the region

$$\Omega = \Big\{ (S(t), A(t), W(t), P(t), D(t)) \in \mathbb{R}^5_+ \;\Big|\; 0 \leq S(t) + A(t) + W(t) + P(t) + D(t) \leq \frac{\Lambda}{\mu} \Big\}.$$

If the initial conditions satisfy

$$S(0), A(0), W(0), P(0), D(0) \geq 0,$$

then the region Ω is positively invariant under the flow of system (1).

Proof. Let

$$N(t) = S(t) + A(t) + W(t) + P(t) + D(t).$$

Summing the equations of system (1) gives

$$\frac{dN(t)}{dt} = \Lambda - \mu N(t).$$

This implies the differential inequality

$$\frac{dN(t)}{dt} \leq \Lambda - \mu N(t).$$

By standard comparison theorems, we have

$$\limsup_{t \to +\infty} N(t) \leq \frac{\Lambda}{\mu}.$$

Thus,

$$N(t) \leq \frac{\Lambda}{\mu} \quad \text{for all } t \geq 0.$$

Therefore, trajectories starting in Ω remain in Ω for all time, proving that Ω is positively invariant for system (1) [14]. □

B.2.2 Existence of solutions

This subsection addresses the fundamental question of the existence and uniqueness of solutions for the compartmental model proposed in this study. To this end, we state and prove the following theorem.

Theorem 3. System (1), with initial conditions

$$S(0) \geq 0, \quad A(0) \geq 0, \quad W(0) \geq 0, \quad P(0) \geq 0, \quad D(0) \geq 0,$$

admits a unique solution.

Proof. Define

$$X(t) = \begin{pmatrix} S(t) \\ A(t) \\ W(t) \\ P(t) \\ D(t) \end{pmatrix}, \quad \Upsilon(X) = KX + B(X),$$

where

$$K = \begin{bmatrix} -(\alpha+\mu) & 0 & 0 & 0 & 0 \\ 0 & -(\gamma+\mu) & 0 & 0 & 0 \\ \alpha & \gamma & -(\eta+\delta+\mu) & \theta & 0 \\ 0 & 0 & \eta & -(\theta+\mu) & 0 \\ 0 & 0 & \delta & 0 & -\mu \end{bmatrix},$$

$$B(X) = \begin{pmatrix} \Lambda - \beta SA \\ \beta SA \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

To establish uniqueness, we show that Υ is Lipschitz continuous. For any $Z, Y \in \mathbb{R}^5$, we have

$$\|B(Z) - B(Y)\| \leq 2|\beta S_Z A_Z - \beta S_Y A_Y|.$$

By applying the triangle inequality, we have

$$\begin{aligned} |\beta S_Z A_Z - \beta S_Y A_Y| &\leq \beta |S_Z(A_Z - A_Y)| + \beta |A_Y(S_Z - S_Y)| \\ &\leq \beta \|S_Z\| \|A_Z - A_Y\| + \beta \|A_Y\| \|S_Z - S_Y\|. \end{aligned}$$

Since all solutions are bounded in the positively invariant region Ω, where

$$S(t), A(t) \leq \frac{\Lambda}{\mu},$$

we conclude that

$$\|B(Z) - B(Y)\| \leq 2\beta \frac{\Lambda}{\mu} \|Z - Y\| = C\|Z - Y\|, \quad \text{with } C = 2\beta \frac{\Lambda}{\mu}.$$

Moreover, since K is a constant matrix, it is also Lipschitz continuous with Lipschitz constant $\|K\|$. Hence,

$$\|\Upsilon(Z) - \Upsilon(Y)\| \leq G\|Z - Y\|, \quad \text{where } G = \max(C, \|K\|) < \infty.$$

Thus, by the Picard–Lindelöf theorem, system (1) admits a unique solution for the given initial data [22]. □

C Stability and sensitivity analysis of the model

To gain deeper insights into the dynamics of the proposed compartmental model, we perform a stability analysis and a sensitivity analysis of the model parameters. These analyses are useful for understanding the system dynamics and identifying the parameters that have the greatest influence on e-commerce refusal.

C.1 Computation of the basic reproduction number

The basic reproduction number R_0 represents the average number of new merchants who become refusers of e-commerce due to one refuser introduced into an otherwise refusal-free population. To compute R_0, we use the next-generation matrix approach introduced by Van den Driessche and Watmough [26].

We reformulate system (1) in the form

$$\frac{d\mathbf{X}}{dt} = \mathcal{F}(\mathbf{X}) - \mathcal{V}(\mathbf{X}),$$

where

$$\mathbf{X} = \begin{pmatrix} A \\ W \\ S \end{pmatrix},$$

and

$$\mathcal{F}(\mathbf{X}) = \begin{pmatrix} \beta SA \\ 0 \\ 0 \end{pmatrix}, \qquad \mathcal{V}(\mathbf{X}) = \begin{pmatrix} (\gamma + \mu)A \\ -\alpha S - \gamma A - \theta P + (\eta + \delta + \mu)W \\ -\Lambda + \beta SA + \alpha S + \mu S \end{pmatrix}.$$

At the **refusal-free equilibrium**

$$E_0 = (A_0, W_0, S_0),$$

where

$$A_0 = 0, \qquad W_0 = 0, \qquad S_0 = \frac{\Lambda}{\alpha + \mu},$$

the Jacobian matrices are given by

$$J_{\mathcal{F}}(E_0) = \begin{pmatrix} \beta S_0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} \frac{\beta\Lambda}{\alpha+\mu} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Hence, the next-generation matrix associated with the three state variables (A, W, S) is

$$F = \begin{pmatrix} \frac{\beta\Lambda}{\alpha+\mu} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The Jacobian of $\mathcal{V}(\mathbf{X})$ evaluated at E_0 is

$$V = J_{\mathcal{V}}(E_0) = \begin{pmatrix} \gamma + \mu & 0 & 0 \\ -\gamma & \eta + \delta + \mu & -\alpha \\ \frac{\beta\Lambda}{\alpha+\mu} & 0 & \alpha + \mu \end{pmatrix}.$$

Its inverse is

$$V^{-1} = \begin{pmatrix} \frac{1}{\gamma+\mu} & 0 & 0 \\ \frac{\gamma}{(\gamma+\mu)(\eta+\delta+\mu)} & \frac{1}{\eta+\delta+\mu} & \frac{\alpha}{(\alpha+\mu)(\eta+\delta+\mu)} \\ -\frac{\beta\Lambda}{(\alpha+\mu)^2(\gamma+\mu)} & 0 & \frac{1}{\alpha+\mu} \end{pmatrix}.$$

Therefore,

$$FV^{-1} = \begin{pmatrix} \frac{\beta\Lambda}{(\alpha+\mu)(\gamma+\mu)} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The eigenvalues of FV^{-1} are

$$\lambda_1 = \frac{\beta\Lambda}{(\alpha+\mu)(\gamma+\mu)}, \qquad \lambda_2 = 0, \qquad \lambda_3 = 0.$$

Consequently, the basic reproduction number, which is defined as the spectral radius of FV^{-1}, is

$$R_0 = \rho(FV^{-1}) = \frac{\beta\Lambda}{(\alpha+\mu)(\gamma+\mu)}.$$

C.2 Equilibrium points analysis

The model admits two equilibrium points. The first, E_0, corresponds to the absence of refusal to use e-commerce, while the second, E_*, corresponds to the presence of refusal to use e-commerce among merchants.

C.3 Refusal-free equilibrium point E_0

The equilibrium E_0 corresponds to a situation in which there is no refusal to use e-commerce among merchants. It is given by

$$E_0 = (S_0, A_0, W_0, P_0, D_0),$$

where

$$\begin{aligned}
S_0 &= \frac{\Lambda}{\alpha+\mu},\\
A_0 &= 0,\\
W_0 &= \frac{\alpha\Lambda(\theta+\mu)}{(\alpha+\mu)(\mu^2+\theta\mu+\theta\delta+\mu\delta+\mu\eta)},\\
P_0 &= \frac{\alpha\Lambda\eta}{(\alpha+\mu)(\mu^2+\theta\mu+\theta\delta+\mu\delta+\mu\eta)},\\
D_0 &= \frac{\alpha\Lambda\delta(\theta+\mu)}{\mu(\alpha+\mu)(\mu^2+\theta\mu+\theta\delta+\mu\delta+\mu\eta)}.
\end{aligned}$$

Since $A_0 = 0$, this equilibrium represents the absence of refusal to use e-commerce among merchants.

C.4 Equilibrium point with refusal of e-commerce E_*

The equilibrium E_* corresponds to a situation in which refusal to use e-commerce is present among merchants. It is given by

$$E_* = (S_*, A_*, W_*, P_*, D_*)\,,$$

with

$$\begin{aligned}
S_* &= \frac{\gamma+\mu}{\beta},\\
A_* &= \frac{(\alpha+\mu)(R_0-1)}{\beta},\\
W_* &= \frac{\theta+\mu}{\beta}\cdot\frac{\alpha\mu+\alpha\gamma R_0+\gamma\mu(R_0-1)}{\mu^2+\theta\mu+\theta\delta+\mu\delta+\mu\eta},\\
P_* &= \frac{\dfrac{\alpha(\gamma+\mu)}{\beta}+\dfrac{\gamma(\alpha+\mu)(R_0-1)}{\beta}}{\mu+\dfrac{\mu(\theta+\mu)}{\eta}+\dfrac{\delta(\theta+\mu)}{\eta}},\\
D_* &= \frac{\delta(\theta+\mu)}{\beta\mu}\cdot\frac{\alpha\mu+\gamma\mu(R_0-1)+\alpha\gamma R_0}{\mu^2+\theta\mu+\theta\delta+\mu\delta+\mu\eta}.
\end{aligned}$$

Since

$$A_* = \frac{(\alpha+\mu)(R_0-1)}{\beta},$$

we have

$$A_* > 0 \qquad \text{if and only if} \qquad R_0 > 1.$$

Thus, when $R_0 < 1$, the refusal of e-commerce cannot persist and the equilibrium E_0 describes the absence of refusal. When $R_0 > 1$, refusal to use e-commerce is present and persists at the equilibrium E_*.

C.5 Local stability analysis

In this subsection, we conduct a local stability analysis of the two equilibrium points. We first consider the refusal-free equilibrium E_0.

Theorem 4. The refusal-free equilibrium point E_0 is locally asymptotically stable if $R_0 < 1$.

Proof. We begin by considering the Jacobian matrix J_0, evaluated at the equilibrium point E_0,

$$J_0(E_0) = \begin{bmatrix} -\alpha-\mu & -\Lambda\frac{\beta}{\alpha+\mu} & 0 & 0 & 0 \\ 0 & (R_0-1)(\gamma+\mu) & 0 & 0 & 0 \\ \alpha & \gamma & -\mu-\delta-\eta & \theta & 0 \\ 0 & 0 & \eta & -\theta-\mu & 0 \\ 0 & 0 & \delta & 0 & -\mu \end{bmatrix}.$$

The eigenvalues of this matrix are given by

$$\begin{aligned} v_1 &= -\frac{1}{2}\theta - \mu - \frac{1}{2}\delta - \frac{1}{2}\eta - \frac{1}{2}\sqrt{(\theta-\delta)^2 + 2\theta\eta + 2\delta\eta + \eta^2} < 0, \\ v_2 &= -\alpha - \mu < 0, \\ v_3 &= -\mu < 0, \\ v_4 &= (R_0-1)(\gamma+\mu) < 0 \quad (\text{since } R_0 < 1), \\ v_5 &= \frac{1}{2}\sqrt{(\theta-\delta)^2 + 2\theta\eta + 2\delta\eta + \eta^2} - \left(\mu + \frac{1}{2}\delta + \frac{1}{2}\eta + \frac{1}{2}\theta\right) < 0. \end{aligned}$$

To demonstrate that $v_5 < 0$, we compute

$$\begin{aligned} &\left(\frac{1}{2}\sqrt{(\theta-\delta)^2 + 2\theta\eta + 2\delta\eta + \eta^2}\right)^2 - \left(\mu + \frac{1}{2}\delta + \frac{1}{2}\eta + \frac{1}{2}\theta\right)^2 \\ &\quad = -\mu^2 - \theta\mu - \theta\delta - \mu\delta - \mu\eta < 0. \end{aligned}$$

Since all eigenvalues are negative, the refusal-free equilibrium point E_0 is locally asymptotically stable if $R_0 < 1$. □

We now consider the equilibrium point E_*, which corresponds to the presence of e-commerce refusal among merchants.

Theorem 5. The equilibrium point E_* with e-commerce refusal is locally asymptotically stable if $R_0 > 1$.

Proof. We have the Jacobian matrix J_* evaluated at E_*:

$$J_*(E_*) = \begin{bmatrix} -R_0(\alpha+\mu) & -(\gamma+\mu) & 0 & 0 & 0 \\ (R_0-1)(\alpha+\mu) & 0 & 0 & 0 & 0 \\ \alpha & \gamma & -(\mu+\delta+\eta) & \theta & 0 \\ 0 & 0 & \eta & -(\theta+\mu) & 0 \\ 0 & 0 & \delta & 0 & -\mu \end{bmatrix}.$$

The characteristic polynomial of $J(E_*)$ is given by

$$P(X) = X^5 + b_1X^4 + b_2X^3 + b_3X^2 + b_4X + b_5,$$

where

$$\begin{aligned}
b_1 =& (\theta + 3\mu + \delta + \eta + R_0(\alpha + \mu)) > 0,\\
b_2 =& 2\mu^2 + 2\theta\mu + \theta\delta + 2\mu\delta + 2\mu\eta + 4\mu^2R_0 + \theta\alpha R_0 + \theta\mu R_0 + 3\alpha\mu R_0\\
&+ \alpha\delta R_0 + \mu\delta R_0 + \alpha\eta R_0 + \mu\eta R_0 + (R_0 - 1)(\alpha\gamma + \alpha\mu + \gamma\mu) > 0,\\
b_3 =& (R_0 - 1)\begin{pmatrix} 2\mu^3 + 3\alpha\mu^2 + 3\gamma\mu^2 + \theta\alpha\gamma + \theta\alpha\mu + \theta\gamma\mu \\ +3\alpha\gamma\mu + \alpha\gamma\delta + \alpha\mu\delta + \gamma\mu\delta + \alpha\gamma\eta + \alpha\mu\eta + \gamma\mu\eta \end{pmatrix}\\
&+ 4\mu^3R_0 + 3\theta\mu^2R_0 + 3\alpha\mu^2R_0 + 3\mu^2\delta R_0 + 3\mu^2\eta R_0\\
&+ \theta\mu\delta + 2\theta\alpha\mu R_0 + \theta\alpha\delta R_0 + \theta\mu\delta R_0 + 2\alpha\mu\delta R_0 + 2\alpha\mu\eta R_0 > 0,\\
b_4 =& -(\alpha + \mu)\begin{pmatrix} -(R_0 - 1)\begin{pmatrix} 4\mu^3 + 3\theta\mu^2 + 3\gamma\mu^2 + 3\mu^2\delta \\ +3\mu^2\eta + 2\theta\gamma\mu + \theta\gamma\delta \\ +2\theta\mu\delta + 2\gamma\mu\delta + 2\gamma\mu\eta \end{pmatrix} \\ -\mu^3 - \theta\mu^2 - \mu^2\delta - \theta\mu\delta - \mu^2\eta \end{pmatrix} > 0,\\
b_5 =& \mu(R_0 - 1)(\alpha + \mu)(\gamma + \mu)\left(\mu^2 + \theta\mu + \theta\delta + \mu\delta + \mu\eta\right) > 0.
\end{aligned}$$

Using the Routh–Hurwitz criterion, the system is locally asymptotically stable if $b_1 > 0$, $b_2 > 0$, $b_3 > 0$, $b_4 > 0$, $b_5 > 0$, $b_1b_2b_3 > b_3^2 + b_1^2b_4$, and

$$(b_1b_4 - b_5)\left(b_1b_2b_3 - b_3^2 + b_1^2b_4\right) > b_5(b_1b_2 - a_3)^2 + b_1b_5^2$$

for $R_0 > 1$.

Finally, if $R_0 > 1$, then the equilibrium point E_* with e-commerce refusal is locally asymptotically stable. □

C.6 Global stability analysis

The objective of this subsection is to examine the global stability of the refusal-free equilibrium E_0 and the equilibrium point E_* associated with e-commerce refusal. The following analysis describes the global stability properties of these two equilibrium points.

Theorem 6. The refusal-free equilibrium E_0 is globally asymptotically stable if $R_0 \leq 1$.

Proof. Let us consider the Lyapunov function $V : \Omega \to \mathbb{R}$, defined as

$$V(S,A) = \frac{1}{2}\left((S - S_0) + A\right)^2 + \frac{\alpha + \gamma + 2\mu}{\beta} A, \tag{2}$$

where S and A represent the state variables, S_0 is a constant, and α, γ, μ, and β are parameters of the system.

The time derivative of $V(S,A)$ is expressed as

$$\dot{V}(S,A) = \left((S - S_0) + A\right)\left(\dot{S} + \dot{A}\right) + \frac{\alpha + \gamma + 2\mu}{\beta}\dot{A}.$$

After performing calculations, we obtain

$$\begin{aligned}\dot{V}(S,A) = &-(\alpha + \mu)\left(S - \frac{\Lambda}{\alpha + \mu}\right)^2 - (\gamma + \mu)A^2 \\ &- (\alpha + \gamma + 2\mu)\frac{(\mu + \gamma)}{\beta}(1 - R_0)A \leq 0.\end{aligned}$$

Thus, $\dot{V}(S,A) = 0 \Longleftrightarrow S = S_0$ and $A = 0$.

Hence, the theorem is proved. Furthermore, by LaSalle's invariance principle [18], the refusal-free equilibrium E_0 is globally asymptotically stable on Ω. □

In the following, we analyze the global stability of the equilibrium point E_* with e-commerce refusal.

Theorem 7. The equilibrium point E_* with e-commerce refusal is globally asymptotically stable if $R_0 > 1$.

Proof. Let us consider the Lyapunov function $V : \Omega \to \mathbb{R}$ defined as

$$V(S,A) = (S - A) - S_* \ln\left(\frac{S}{S_*}\right) + (A - A_*) - A_* \ln\left(\frac{A}{A_*}\right), \tag{3}$$

where S and A are the state variables, and S_* and A_* represent the values of S and A at the equilibrium point E_*.

Taking the time derivative of $V(S,A)$, we obtain

$$\dot{V}(S,A) = \dot{S} + \dot{A} - S_*\frac{\dot{S}}{S} - A_*\frac{\dot{A}}{A}.$$

Further calculations yield

$$\dot{V}(S,A) = -\frac{1}{\beta}\frac{\Lambda}{S(\gamma + \mu)}\left(\gamma + \mu - \beta S\right)^2 \leq 0,$$

which shows that the Lyapunov function is nonincreasing.

Additionally, we find that

$$\dot{V}(S, A) = 0 \Longleftrightarrow S = S_*.$$

Therefore, by LaSalle's invariance principle [18], the equilibrium point E_* with e-commerce refusal is globally asymptotically stable on Ω. □

C.7 Numerical simulation of the model

In this subsection, we present numerical simulations of system (1) to illustrate its dynamical behavior under different parameter settings. The following parameter values are considered:

$$\Lambda = 400{,}000, \quad \mu = 0.39, \quad \alpha = 0.5, \quad \eta = 0.3,$$
$$\gamma = 0.2, \quad \delta = 0.3, \quad \theta = 0.1, \quad \beta = 1 \times 10^{-7}.$$

For these parameter values, the basic reproduction number is

$$R_0 = 0.7618 < 1.$$

Accordingly, the system approaches the refusal-free equilibrium point

$$E_0 = \left(4.4944 \times 10^5,\ 0,\ 2.4195 \times 10^5,\ 1.4813 \times 10^5,\ 1.8612 \times 10^5\right).$$

As shown in Figure 2, the state variables S, A, W, P, and D converge toward the corresponding components of E_0. In particular, the refusal class A decreases toward zero, indicating the disappearance of e-commerce refusal in the long term.

Next, we consider a second scenario in which the transmission rate is increased to

$$\beta = 1.8 \times 10^{-6},$$

while all the other parameter values remain unchanged. In this case, the basic reproduction number satisfies

$$R_0 = 1.33 > 1.$$

The system then approaches the endemic equilibrium point E_* with e-commerce refusal:

$$E_* = \left(3.2778 \times 10^5,\ 1.8352 \times 10^5,\ 2.1598 \times 10^5,\ 1.3223 \times 10^5,\ 1.6614 \times 10^5\right).$$

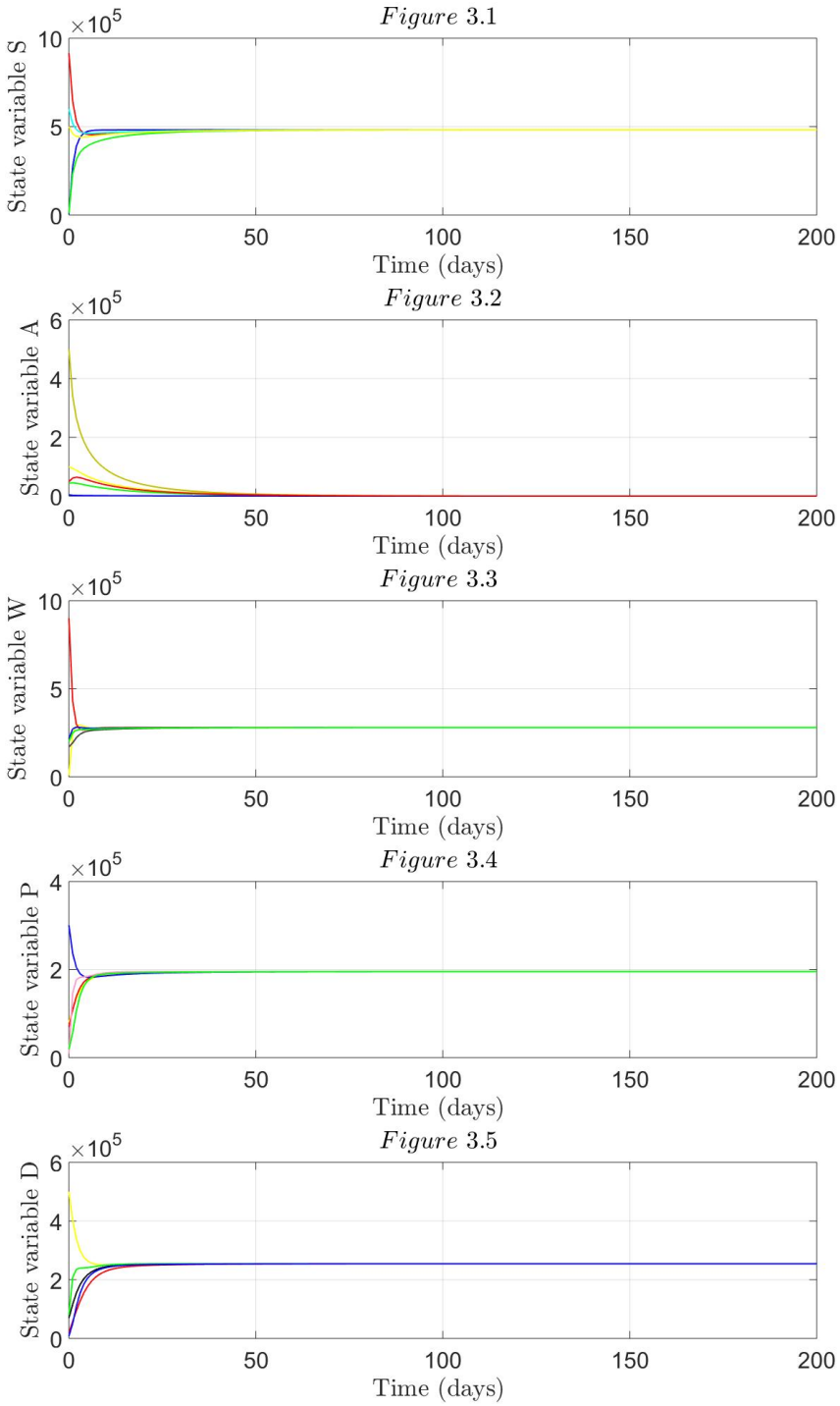


Figure 2: Dynamics converging to the refusal-free equilibrium for e-commerce.

As illustrated in Figure 3, all state variables converge toward the corresponding components of E_*. In particular, the refusal class A converges to a positive value, indicating that e-commerce refusal persists in the merchant population over time.

These numerical simulations illustrate the influence of the basic reproduction number R_0 on the long-term dynamics of the model. When $R_0 < 1$, e-commerce refusal decreases over time and the system converges to the refusal-free equilibrium E_0. In contrast, when $R_0 > 1$, the refusal class persists and the system approaches the equilibrium E_*, where the number of refusing merchants remains positive. These results emphasize the importance of controlling the mechanisms that contribute to the transmission of e-commerce refusal in order to promote e-commerce adoption among merchants.

C.8 Sensitivity analysis of model parameters

Sensitivity analysis is a key tool used to evaluate how variations in model parameters influence the system's output. In particular, it allows us to identify which parameters have the greatest impact on the basic reproduction number R_0, which governs the spread of e-commerce refusal among merchants [3]. In this subsection, we assess the effect of each parameter in system (1) on R_0.

Definition

The normalized forward sensitivity index of a variable v with respect to a parameter p is defined by

$$\Delta_p^v = \frac{\partial v}{\partial p} \cdot \frac{p}{v}.$$

For our model, the basic reproduction number is given by

$$R_0 = \beta \cdot \frac{\Lambda}{(\alpha+\mu)(\gamma+\mu)}.$$

Computation of sensitivity indices

Applying the definition, the sensitivity indices of R_0 with respect to each parameter are

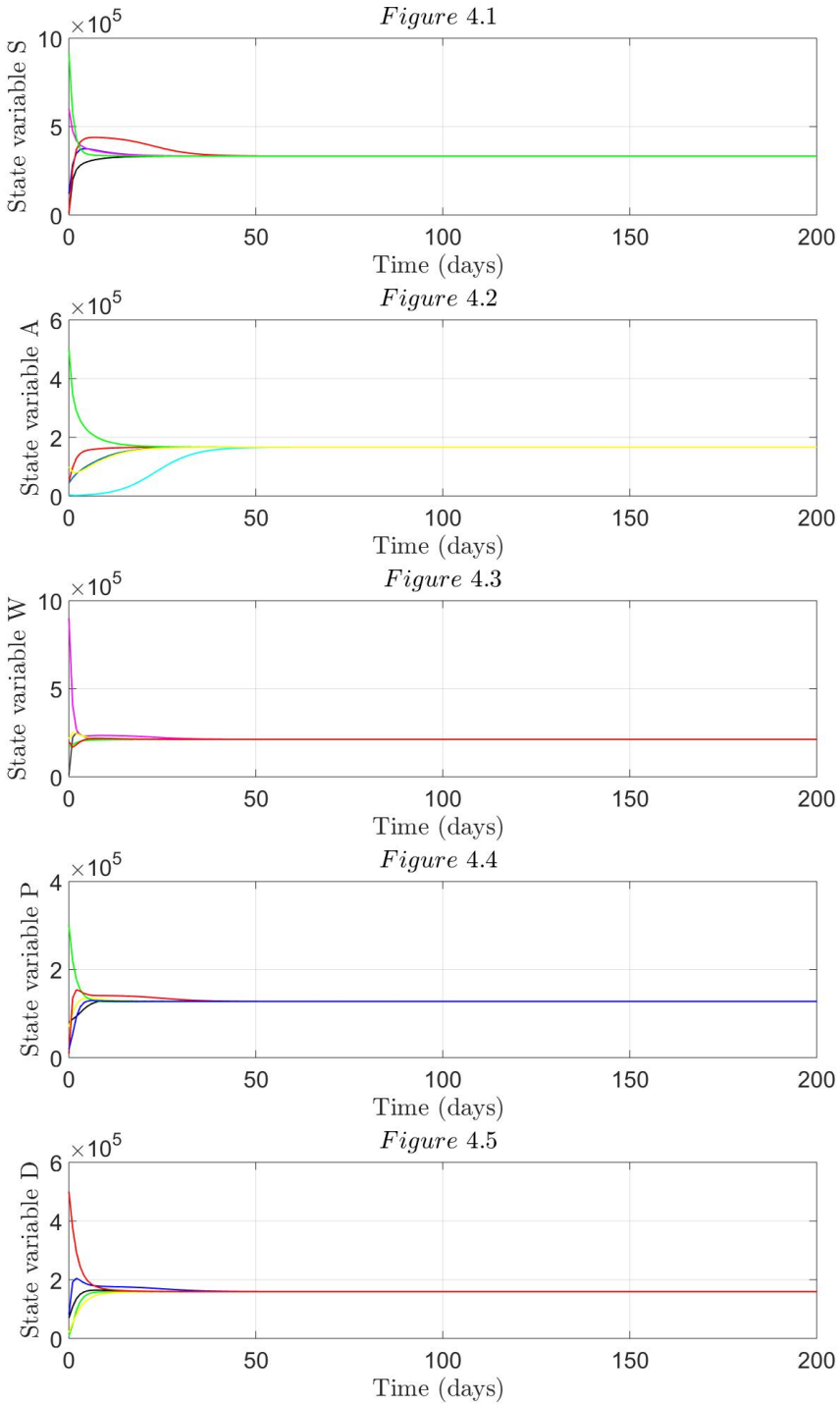


Figure 3: Dynamics converging to the equilibrium with e-commerce refusal.

$$\begin{cases} \Delta_{\Lambda}^{R_0} = 1, \\ \Delta_{\beta}^{R_0} = 1, \\ \Delta_{\alpha}^{R_0} = -\dfrac{\alpha}{\alpha+\mu} < 0, \\ \Delta_{\gamma}^{R_0} = -\dfrac{\gamma}{\gamma+\mu} < 0, \\ \Delta_{\mu}^{R_0} = -\dfrac{\mu(\alpha+\gamma+2\mu)}{(\alpha+\mu)(\gamma+\mu)} < 0. \end{cases} \tag{4}$$

Interpretation

The sign of each sensitivity index indicates the direction of its effect on R_0. A positive sensitivity index, such as $\Delta_{\Lambda}^{R_0} = 1$, means that increasing the corresponding parameter increases R_0. For instance, a 10% increase in β leads to a 10% increase in R_0.

In contrast, a negative sensitivity index, such as $\Delta_{\alpha}^{R_0} < 0$, implies that increasing the corresponding parameter reduces R_0. For example, raising α by 10% reduces R_0 proportionally to

$$\frac{10\% \cdot \alpha}{\alpha+\mu}.$$

The sensitivity indices are illustrated in Figure 4, which provides a graphical representation of the influence of the model parameters on R_0.

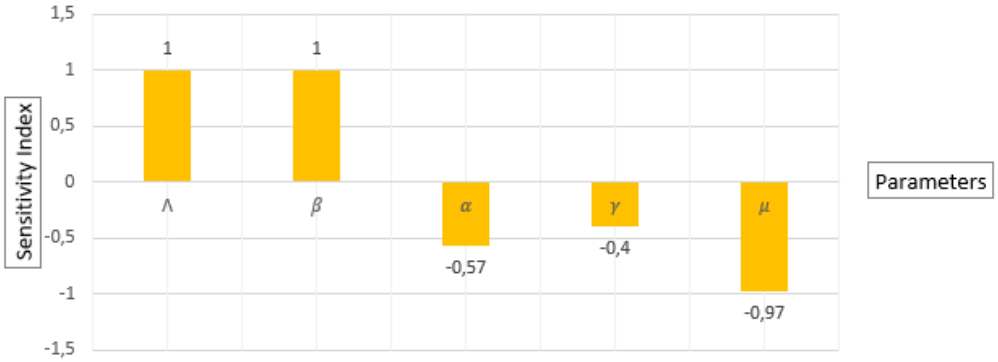


Figure 4: Sensitivity indices of R_0 with respect to model parameters.

Remark 1. The sensitivity analysis shows that the recruitment rate Λ and the rejection influence rate β increase R_0, thereby promoting the persistence of e-commerce refusal. In contrast,

increasing either the willingness rate α or the recovery rate γ decreases R_0, thereby contributing to the reduction of e-commerce refusal. The reintegration rate θ does not directly affect R_0, since it does not appear in its expression.

These findings motivate the introduction of three time-dependent control functions:

- $u_1(t)$: awareness campaigns,
- $u_2(t)$: cybersecurity support,
- $u_3(t)$: technical assistance.

By targeting the main factors and processes influencing merchant behavior, these control strategies provide a practical approach to reducing e-commerce refusal and promoting digital adoption among merchants.

D Formulation of the optimal control problem

In this section, we formulate an optimal control problem aimed at increasing the number of merchants who utilize e-commerce and reducing those who refrain from its use. The proposed control strategies are incorporated into a mathematical model to achieve this objective.

D.1 Mathematical model with control variables

E-commerce enables online store owners to showcase their products and services to customers at any time of the day or night, throughout the week. Furthermore, they can conduct transactions both locally and internationally. Moreover, it allows them to communicate with a diverse customer base, facilitated by the ongoing global digital revolution, which increases customer reach.

To augment the number of merchants who utilize e-commerce to fulfill their daily requirements and curtail the number of those who eschew its use, we introduce three control variables:

- $u_1(t)$: Advertising campaigns across all visual and audio communication channels, including social media, to encourage stores to adopt e-commerce and motivate merchants to utilize it.

- $u_2(t)$: Provision of insurance for all shops against potential cybercrime exposure and creation of safer websites for commercial shops to protect against cybercrime, thereby motivating merchants to use e-commerce.

- $u_3(t)$: Establishment of dedicated teams to train and educate merchants on how to manage e-commerce sites, especially concerning internet and maintenance issues.

By incorporating these control variables into the initial system (1), we obtain the following controlled dynamical system:

$$\begin{cases} \dfrac{dS(t)}{dt} = \Lambda - \beta S(t)A(t) - \alpha S(t) - \mu S(t) - u_1(t)S(t), \\ \dfrac{dA(t)}{dt} = \beta S(t)A(t) - \gamma A(t) - \mu A(t) - u_2(t)A(t), \\ \dfrac{dW(t)}{dt} = \alpha S(t) + \gamma A(t) + \theta P(t) - (\eta + \delta + \mu)W(t) \\ \qquad\qquad + u_1(t)S(t) + u_2(t)A(t) + u_3(t)P(t), \\ \dfrac{dP(t)}{dt} = \eta W(t) - \theta P(t) - \mu P(t) - u_3(t)P(t), \\ \dfrac{dD(t)}{dt} = \delta W(t) - \mu D(t). \end{cases} \tag{5}$$

The initial conditions are given by

$$S(0) \geq 0, \quad A(0) \geq 0, \quad W(0) \geq 0, \quad P(0) \geq 0, \quad D(0) \geq 0.$$

D.2 Objective functional

Our aim is to minimize the total number of merchants who refuse to use e-commerce over a fixed time horizon $[t_0, t_f]$. This is achieved by optimizing the control variables $u_1(t)$, $u_2(t)$, and $u_3(t)$ in the model (5). The objective functional to be minimized is defined as

$$\begin{aligned} &J(u_1, u_2, u_3) \\ &\quad = A(t_f) + P(t_f) - W(t_f) \\ &\qquad + \int_{t_0}^{t_f} \left[A(t) + P(t) - W(t) + \frac{C_1}{2}u_1^2(t) + \frac{C_2}{2}u_2^2(t) + \frac{C_3}{2}u_3^2(t) \right] dt, \end{aligned} \tag{6}$$

where $C_1, C_2, C_3 > 0$ are weight parameters representing the cost associated with implementing each control strategy.

D.3 Admissible control set

We seek optimal controls $u_1^*(t)$, $u_2^*(t)$, and $u_3^*(t)$ such that

$$J(u_1^*, u_2^*, u_3^*) = \min_{(u_1,u_2,u_3)\in U} J(u_1, u_2, u_3), \tag{7}$$

where the set of admissible controls U is defined by

$$\begin{aligned} U = \{(u_1(t), u_2(t), u_3(t)) \,|\, & 0 \le u_1(t) \le 1,\ 0 \le u_2(t) \le 1, \\ & 0 \le u_3(t) \le 1,\ \text{for } t \in [t_0, t_f]\}\,. \end{aligned} \tag{8}$$

D.4 Problem statement

The optimal control problem can be formally stated as follows.

Problem

Find control functions $u_1^*(t)$, $u_2^*(t)$, and $u_3^*(t)$ belonging to the admissible control set U that minimize the objective functional $J(u_1, u_2, u_3)$ given by (6), subject to the controlled system dynamics (5) and initial conditions.

The solution to this optimal control problem will provide the optimal implementation of the proposed strategies to maximize the adoption of e-commerce among merchants while minimizing the associated costs.

D.5 Existence of an optimal control solution

This subsection establishes the existence of an optimal control (u_1^*, u_2^*, u_3^*) that minimizes the cost functional J. The proof is formulated to meet the standards expected by specialists in control theory.

Theorem 8. An optimal control $(u_1^*, u_2^*, u_3^*) \in U$ exists such that

$$J(u_1^*, u_2^*, u_3^*) = \inf_{(u_1,u_2,u_3)\in U} J(u_1, u_2, u_3).$$

Proof. To demonstrate the existence of an optimal control, we utilize standard results from the theory of optimal control as outlined in [12]. The following conditions must be satisfied:

(C1) The set of admissible controls U and the associated state trajectories are nonempty.

(C2) The control set U is convex and closed.

(C3) The right-hand side of the state system is continuous and satisfies linear growth with respect to the state and control variables.

(C4) The cost functional $J(u_1, u_2, u_3)$ is convex with respect to the controls and lower semi-continuous.

Verification of Condition C1: The nonemptiness of U and the existence of corresponding state trajectories follow from the well-posedness of the control system (5), as established in [20].

Verification of Condition C2: By definition, the control set U consists of all measurable functions $(u_1(t), u_2(t), u_3(t))$ such that $0 \le u_i(t) \le 1$ for all $t \in [t_0, t_f]$ and $i = 1, 2, 3$. This set is convex and closed in the space of measurable functions.

Verification of Condition C3: The right-hand side of the system (5) is linear in the control variables and satisfies Lipschitz continuity with respect to the state variables. Additionally, the state variables are known to be bounded due to the modeling and practical considerations inherent in the model, as discussed in the analysis of system (1).

Verification of Condition C4: The cost functional comprises a linear combination of the state variables and quadratic terms in the control variables:

$$\begin{aligned} &J(u_1, u_2, u_3) \\ &= \phi\big(S(t_f), A(t_f), W(t_f), P(t_f), D(t_f)\big) \\ &\quad + \int_{t_0}^{t_f} \left[L\big(A(t), W(t), P(t)\big) + \frac{C_1}{2} u_1^2(t) + \frac{C_2}{2} u_2^2(t) + \frac{C_3}{2} u_3^2(t) \right] dt, \end{aligned}$$

where ϕ and L are linear functions of the state variables. The integrand is convex in the controls due to the quadratic terms, and the functional is lower semicontinuous with respect to the weak convergence of controls in $L^2([t_0, t_f])$.

Having verified all necessary conditions, we invoke the standard existence theorem (see [12]) to conclude that an optimal control (u_1^*, u_2^*, u_3^*) exists in U that minimizes the cost functional J. □

D.6 Characterization of the optimal controls

In this subsection, we derive the necessary conditions for optimality using Pontryagin's maximum principle [23]. By applying this principle to the control system (5) and the cost functional (6), we transform the problem into minimizing the Hamiltonian $\mathcal{H}$ with respect to the control variables u_1, u_2, and u_3.

The Hamiltonian $\mathcal{H}$ is defined as

$$\mathcal{H} = A(t) + P(t) - W(t) + \frac{C_1}{2}u_1^2(t) + \frac{C_2}{2}u_2^2(t) + \frac{C_3}{2}u_3^2(t) + \sum_{i=1}^{5}\lambda_i(t)f_i(t),$$

where $\lambda_i(t)$ are the adjoint variables and $f_i(t)$ are the right-hand sides of the state equations in (5),

$$\begin{cases} f_1(t) = \Lambda - \beta S(t)A(t) - \alpha S(t) - \mu S(t) - u_1(t)S(t), \\ f_2(t) = \beta S(t)A(t) - \gamma A(t) - \mu A(t) - u_2(t)A(t), \\ f_3(t) = \alpha S(t) + \gamma A(t) + \theta P(t) - (\eta + \delta + \mu)W(t) \\ \qquad\quad + u_1(t)S(t) + u_2(t)A(t) + u_3(t)P(t), \\ f_4(t) = \eta W(t) - \theta P(t) - \mu P(t) - u_3(t)P(t), \\ f_5(t) = \delta W(t) - \mu D(t). \end{cases}$$

Theorem 9. Given the optimal controls (u_1^*, u_2^*, u_3^*) and the corresponding state trajectories $S^*(t)$, $A^*(t)$, $W^*(t)$, $P^*(t)$, and $D^*(t)$, there exist absolutely continuous adjoint variables $\lambda_1(t)$, $\lambda_2(t)$, $\lambda_3(t)$, $\lambda_4(t)$, and $\lambda_5(t)$ satisfying the following adjoint equations almost everywhere on $[t_0, t_f]$:

$$\begin{cases} \dfrac{d\lambda_1}{dt} = \lambda_1 \left(\beta A^*(t) + \alpha + \mu + u_1^*(t)\right) - \lambda_2 \beta A^*(t) - \lambda_3 \left(\alpha + u_1^*(t)\right), \\ \dfrac{d\lambda_2}{dt} = 1 + \lambda_1 \beta S^*(t) + \lambda_2 \left(\beta S^*(t) + \gamma + \mu + u_2^*(t)\right) - \lambda_3 \left(\gamma + u_2^*(t)\right), \\ \dfrac{d\lambda_3}{dt} = -1 + \lambda_3 \left(\eta + \delta + \mu\right) - \lambda_4 \eta - \lambda_5 \delta, \\ \dfrac{d\lambda_4}{dt} = 1 + \lambda_3 \left(\theta + u_3^*(t)\right) + \lambda_4 \left(\theta + \mu + u_3^*(t)\right), \\ \dfrac{d\lambda_5}{dt} = \lambda_5 \mu. \end{cases} \tag{9}$$

The transversality conditions at the final time t_f are

$$\begin{cases} \lambda_1(t_f) = 0, \\ \lambda_2(t_f) = 1, \\ \lambda_3(t_f) = -1, \\ \lambda_4(t_f) = 1, \\ \lambda_5(t_f) = 0. \end{cases} \tag{10}$$

Moreover, the optimal controls are given by

$$\begin{cases} u_1^*(t) = \min\left\{1, \max\left\{0, \dfrac{\left(\lambda_1(t) - \lambda_3(t)\right) S^*(t)}{C_1}\right\}\right\}, \\ u_2^*(t) = \min\left\{1, \max\left\{0, \dfrac{\left(\lambda_2(t) - \lambda_3(t)\right) A^*(t)}{C_2}\right\}\right\}, \\ u_3^*(t) = \min\left\{1, \max\left\{0, \dfrac{\left(\lambda_4(t) - \lambda_3(t)\right) P^*(t)}{C_3}\right\}\right\}. \end{cases} \tag{11}$$

Proof. Applying Pontryagin's maximum principle, the Hamiltonian $\mathcal{H}$ is constructed as given above. The adjoint equations are derived by taking the negative partial derivatives of the Hamiltonian with respect to the state variables. Specifically, for $i = 1, \ldots, 5$,

$$\frac{d\lambda_i}{dt} = -\frac{\partial \mathcal{H}}{\partial x_i},$$

where $x_1 = S$, $x_2 = A$, $x_3 = W$, $x_4 = P$, and $x_5 = D$.

Calculating these derivatives yields the adjoint equations (9). The transversality conditions (10) are obtained from the terminal conditions of the cost functional.

The optimality conditions require that the Hamiltonian be minimized with respect to the control variables almost everywhere on $[t_0, t_f]$. Taking the partial derivatives of $\mathcal{H}$ with respect to u_1, u_2, and u_3 and setting them to zero gives

$$\begin{cases} \dfrac{\partial \mathcal{H}}{\partial u_1} = C_1 u_1^*(t) - \lambda_1(t) S^*(t) + \lambda_3(t) S^*(t) = 0, \\ \dfrac{\partial \mathcal{H}}{\partial u_2} = C_2 u_2^*(t) - \lambda_2(t) A^*(t) + \lambda_3(t) A^*(t) = 0, \\ \dfrac{\partial \mathcal{H}}{\partial u_3} = C_3 u_3^*(t) - \lambda_3(t) P^*(t) + \lambda_4(t) P^*(t) = 0. \end{cases}$$

Solving for the controls, we obtain expressions for $u_1^*(t)$, $u_2^*(t)$, and $u_3^*(t)$ as given in (11). These expressions are then projected onto the admissible control set U by applying the bounds $0 \leq u_i(t) \leq 1$.

□

E Numerical simulations and discussion

In this section, we present numerical simulations of the optimal control strategies implemented in the proposed model using MATLAB. The simulations aim to illustrate the dynamic effects of the control interventions on the spread and mitigation of e-commerce rejection among merchants.

Initial conditions:

$$S(0) = 120{,}000, \quad A(0) = 100{,}000, \quad W(0) = 200{,}000,$$
$$P(0) = 80{,}000, \quad D(0) = 500{,}000.$$

Parameter values used:

$$\Lambda = 400{,}000, \quad \mu = 0.39, \quad \alpha = 0.5, \quad \eta = 0.3,$$

$$\gamma = 0.2, \quad \delta = 0.3, \quad \theta = 0.1, \quad \beta = 1.8 \times 10^{-7}.$$

These parameter settings ensure that the basic reproduction number R_0 exceeds one, indicating the persistence of e-commerce rejection in the absence of control. The upcoming simulation results analyze how appropriately designed control policies can reduce resistance and promote digital adoption across the merchant population.

E.1 Strategy u_1: Promotion through advertising

Figure 5 illustrates the effect of implementing the control strategy $u_1(t)$, which focuses on promoting the importance of e-commerce through advertising. The figure compares the dynamics of the active e-commerce merchant group $W(t)$ with and without the control u_1. It is evident that the application of u_1 leads to a significant increase in the size of the group $W(t)$ over time. This demonstrates the positive impact of advertising on encouraging merchants to adopt e-commerce.

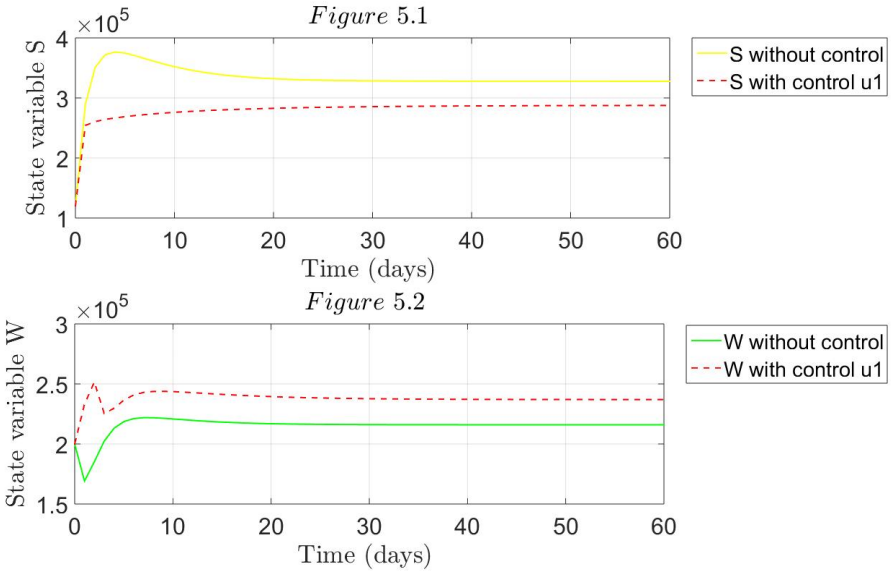


Figure 5: Dynamics of $W(t)$ with and without control u_1.

E.2 Strategy u_2: Cybercrime insurance provision

In Figure 6, we examine the effects of the control strategy $u_2(t)$, which involves providing cybercrime insurance to merchants. The figure shows the changes in the size of the willing merchant group $W(t)$ compared to the scenario without control u_2. The results indicate that implementing u_2 increases the size of $W(t)$ over time, while the size of the e-commerce-refusal merchant group $A(t)$ decreases. This suggests that offering cybercrime insurance effectively encourages merchants to transition from being merely aware to actively willing to use e-commerce.

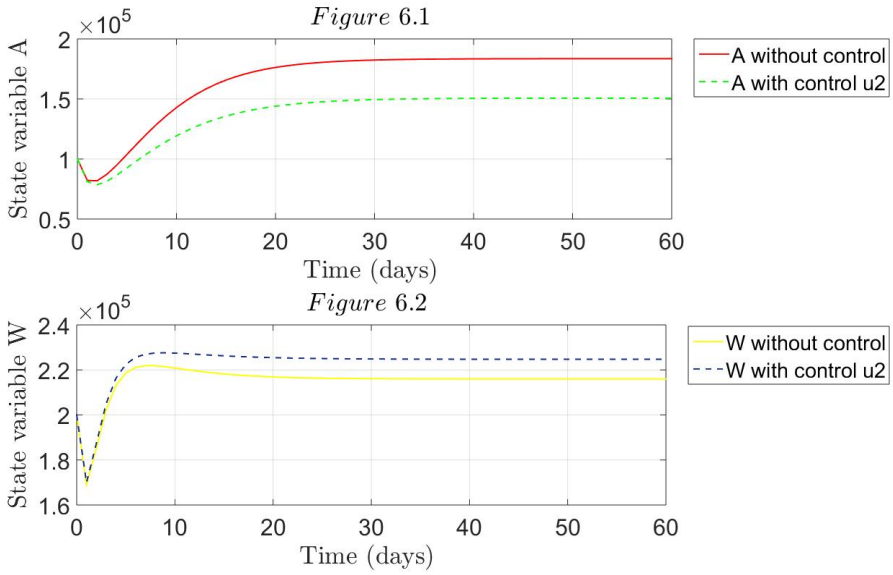


Figure 6: Dynamics of $W(t)$ and $A(t)$ with and without control u_2.

E.3 Strategy u_3: Training and support services

Figure 7 depicts the impact of the control strategy $u_3(t)$, which focuses on providing training and support services to merchants. The figure compares the dynamics of the active e-commerce merchant group $W(t)$ with and without the control u_3. The application of u_3 results in an increase in $W(t)$ over time, while the size of the temporarily inactive merchants group $P(t)$ decreases. This indicates that offering training and support accelerates the transition of merchants from being in the process of adopting e-commerce to becoming willing users.

E.4 Combined Strategies u_1, u_2, and u_3

Finally, Figure 8 shows the effects of implementing all three control strategies simultaneously. The figure compares the dynamics of the willing merchant group $W(t)$ with and without the combined controls. The results demonstrate a significant increase in $W(t)$ when all controls are applied. Concurrently, the sizes of the susceptible group $S(t)$, the e-commerce-refusal group $A(t)$, and the group of merchants in process $P(t)$ decrease over time. This suggests that the combined application of advertising, cybercrime insurance, and training services has a synergistic effect, greatly enhancing the adoption of e-commerce among merchants.

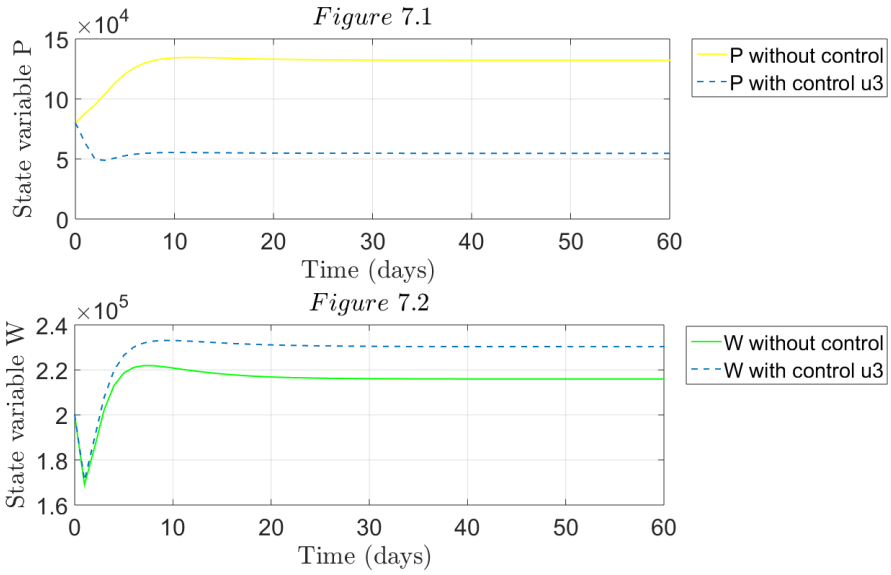


Figure 7: Dynamics of $W(t)$ and $P(t)$ with and without control u_3.

E.5 Cost-effectiveness analysis

To evaluate the efficiency of various intervention strategies aimed at increasing e-commerce adoption among merchants, we employ the average cost-effectiveness ratio (ACER) as a comparative metric [5]. The ACER quantifies the average cost required to prevent a single instance of e-commerce refusal; that is, to stop a merchant from entering or remaining in the refusal compartment A.

Mathematically, the ACER is defined by

$$\text{ACER} = \frac{T_c}{T_a}, \tag{12}$$

where

- T_c denotes the total cost of implementing the intervention strategy over the time horizon $[0, T]$,
- T_a represents the total number of avoided refusals, calculated as the difference in the cumulative number of merchants in compartment A between the uncontrolled and controlled scenarios.

The number of refusals avoided, T_a, is given by

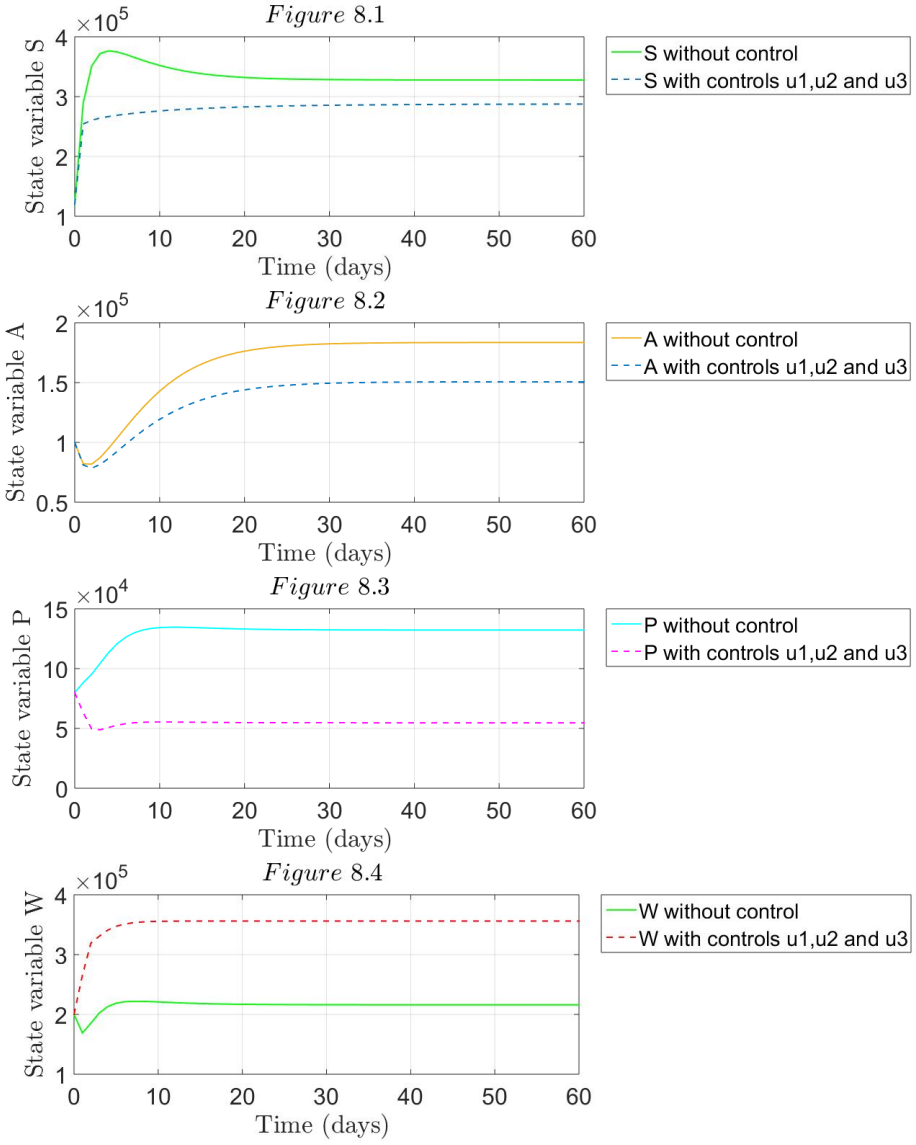


Figure 8: Dynamics of $W(t)$, $S(t)$, $A(t)$, and $P(t)$ with and without controls u_1, u_2, and u_3.

$$T_a = \int_0^T A_{\text{without}}(t)\,dt - \int_0^T A_{\text{with}}(t)\,dt,$$

where $A(t)$ denotes the number of merchants in the refusal compartment at time t, with and without intervention, respectively.

The total intervention cost T_c, assuming constant control efforts, is expressed as

$$T_c = \int_0^T \left(C_1 u_1^2(t) + C_2 u_2^2(t) + C_3 u_3^2(t)\right) dt,$$

where $u_1(t)$, $u_2(t)$, and $u_3(t)$ represent awareness campaigns, cybersecurity reinforcement, and technical training efforts, respectively. The unit costs of these interventions are given by $C_1 = 6$, $C_2 = 8$, and $C_3 = 9$.

Numerical results and interpretation

Table 2 presents the cost-effectiveness outcomes for four different intervention scenarios, each employing constant control values throughout the simulation period. The table reports the total number of avoided refusals T_a, the total cost T_c, and the corresponding ACER for each strategy.

Table 2: Cost-effectiveness analysis results for different scenarios

Scenario	T_a **(avoided refusals)**	T_c **(cost \$)**
Scenario 1	1.7004×10^5	28.80
Scenario 2	3.6109×10^5	86.40
Scenario 4	3.6617×10^5	321.00
Scenario 3	2.5455×10^2	205.80

The following observations can be made:

- **Scenario 1** is the most cost-effective, with an ACER of

$$\text{ACER}_{\text{S1}} = \frac{28.80}{1.7004 \times 10^5} \approx 0.000169,$$

  indicating that a relatively small investment can yield a significant reduction in refusals.

- **Scenario 2** achieves the highest reduction in refusals while maintaining a low ACER,

$$\text{ACER}_{\text{S2}} = \frac{86.40}{3.6109 \times 10^5} \approx 0.000239,$$

  which demonstrates a strong trade-off between cost and impact.

- **Scenario 4** is effective in reducing refusals but incurs a significantly higher cost,

$$\mathrm{ACER}_{\mathrm{S4}} = \frac{321.00}{3.6617 \times 10^5} \approx 0.000877,$$

  making it less cost-efficient compared to Scenarios 1 and 2.

- **Scenario 3** is the least cost-effective, with a high ACER,

$$\mathrm{ACER}_{\mathrm{S3}} = \frac{205.80}{2.5455 \times 10^2} \approx 0.8085,$$

  indicating substantial cost for very limited impact.

Remark 2. Among the evaluated strategies, **Scenario 1** proves to be the most cost-efficient, making it especially valuable in resource-constrained environments. Conversely, for objectives centered on maximizing the total number of refusals avoided, **Scenario 2** delivers the most substantial outcome, while still preserving a favorable cost-effectiveness ratio. These findings offer crucial guidance for decision-makers in optimizing intervention design and resource allocation.

E.6 Discussion

The numerical simulations demonstrate the effectiveness of the proposed control strategies in fostering e-commerce adoption among merchants. Each intervention individually contributes positively to increasing the number of merchants willing to adopt e-commerce.

- **Advertising** (u_1) plays a crucial role in raising awareness and motivating merchants to consider e-commerce.

- **Cybercrime insurance** (u_2) addresses security-related concerns, encouraging merchants to progress from awareness to a genuine willingness to adopt.

- **Training and support** (u_3) helps overcome technical and operational barriers, facilitating the transition toward active adoption.

When combined, these control measures yield a synergistic effect, significantly enhancing adoption rates compared to individual strategies. The observed reduction in the sizes of the susceptible, aware, and in-process merchant compartments indicates that more merchants are advancing swiftly toward active e-commerce participation.

These results highlight the importance of implementing comprehensive and integrated policies that simultaneously raise awareness, mitigate risks, and provide technical support. Such a multifaceted approach is essential for addressing the diverse concerns and obstacles faced by merchants, thereby substantially increasing e-commerce adoption and generating considerable economic benefits.

F Conclusion

In this paper, we developed a mathematical model to analyze the dynamics of e-commerce adoption among merchants, accounting for both adoption and resistance behaviors. The model considers different merchant states, including traditional merchants, merchants refusing e-commerce, active e-commerce merchants, temporarily inactive merchants, and merchants leaving the market. Using the next-generation matrix method, we derived the basic reproduction number R_0 and identified two equilibrium points: The rejection-free equilibrium E_0 and the persistent-refusal equilibrium E_*. The global stability analysis showed that E_0 is globally asymptotically stable when $R_0 \leq 1$, whereas E_* is globally asymptotically stable when $R_0 > 1$, under the corresponding feasibility conditions. Thus, $R_0 = 1$ represents the threshold between the disappearance and persistence of refusal behavior. From a behavioral perspective, $R_0 < 1$ indicates that refusal tends to disappear over time, whereas $R_0 > 1$ allows refusal to persist within the merchant population.

We further formulated the model as an optimal control problem to reduce merchant resistance and promote e-commerce adoption. Using Pontryagin's maximum principle, we derived the necessary optimality conditions and characterized the optimal control strategies.

Numerical simulations confirmed the theoretical results and demonstrated the effectiveness of the proposed interventions. In particular, the combined application of advertising, cybercrime insurance, and training support was the most effective strategy among those considered, increasing e-commerce adoption while reducing merchant resistance. The cost-effectiveness analysis also provided a comparison of the intervention strategies and identified efficient approaches for promoting adoption.

The sensitivity analysis further identified the parameters with the greatest influence on merchant resistance and e-commerce adoption, highlighting the importance of awareness, training, and cybersecurity measures. These findings provide useful guidance for policymakers and stakeholders in designing targeted interventions.

Overall, the proposed framework integrates equilibrium and global stability analysis, sensitivity analysis, optimal control, and cost-effectiveness analysis to provide a quantitative approach for understanding merchant behavior and evaluating e-commerce interventions.

Nevertheless, the model relies on simplifying assumptions, such as population homogeneity and deterministic dynamics. Future work could incorporate merchant heterogeneity, network effects, stochastic dynamics, time-dependent parameters, and empirical data for calibration and validation, thereby improving the practical applicability of the model.

Author Contributions

All authors have made equal contributions to the writing of this article. The final manuscript was read and approved by all authors.

Availability of Data

Data are available upon request.

Acknowledgements

The research reported in this paper was supported by the Moroccan Systems Theory Network.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Ahi, A.A., Sinkovics, N. and Sinkovics, R.R., *E-commerce policy and the global economy: A path to more inclusive development*, Manag. Int. Rev. 63(1) (2023), 27–56.

[2] Balatif, O., Labzai, A. and Rachik, M., *A discrete mathematical modeling and optimal control of the electoral behavior with regard to a political party*, Discrete Dyn. Nat. Soc. 2018 (2018), 1–14.

[3] Balatif, O., Khajji, B. and Rachik, M., *Mathematical modeling, analysis, and optimal control of abstinence behavior of registration on the electoral lists*, Discrete Dyn. Nat. Soc. 2020 (2020) 9738934.

[4] Benfatah, Y., Khaloufi, I., Boutayeb, H., Rachik, M. and Laarabi, H., *Optimal control for a discrete time epidemic model with zones evolution*, Commun. Math. Biol. Neurosci. 51 (2022) 2022.

[5] Berhe, H.W., Gebremeskel, A.A., Melese, Z.T., Al-arydah, M.T. and Gebremichael, A.A., *Modeling and global stability analysis of COVID-19 dynamics with optimal control and cost-effectiveness analysis*, Partial Differ. Equ. Appl. Math. 11 (2024), 100843.

[6] Chen, D.N., Jeng, B., Lee, W.P. and Chuang, C.H., *An agent-based model for consumer-to-business electronic commerce*, Expert Syst. Appl. 34(1) (2008), 469–481.

[7] El Bhih, A., Benfatah, Y. and Rachik, M., *Exact determinations of maximal output admissible set for a class of semilinear discrete systems*, Arch. Control Sci. 30(3) (2020), 523–552.

[8] El Bhih, A., Benfatah, Y., Kouidere, A. and Rachik, M., *A discrete mathematical modeling of transmission of COVID-19 pandemic using optimal control*, Commun. Math. Biol. Neurosci. 2020 (2020).

[9] El Bhih, A., Benfatah, Y., Ghazaoui, A. and Rachik, M., *On the maximal output set of fractional-order discrete-time linear systems*, Math. Model. Comput. 9(2) (2022), 262–277.

[10] El Bhih, A., Benfatah, Y., Hassouni, H., Balatif, O. and Rachik, M., *Mathematical modeling, sensitivity analysis, and optimal control of students awareness in mathematics education*, Partial Differ. Equ. Appl. Math. 11 (2024), 100795.

[11] El Bhih, A., Yaagoub, Z., Rachik, M., et al., *Controlling the dissemination of rumors and antirumors in social networks: A mathematical modeling and analysis approach*, Eur. Phys. J. Plus 139 (2024), 118.

[12] Fleming, W.H. and Rishel, R.W., *Deterministic and Stochastic Optimal Control*, Vol. 1, Springer Science and Business Media, 2012.

[13] Hassouni, H., El Bhih, A. and Balatif, O., *Mathematical modeling, analysis, and optimal control of hacker behavior in cybersecurity systems*, J. Appl. Math. Comput. 72(1) (2026), 32.

[14] Hethcote, H.W. and Thieme, H.R., *Stability of the endemic equilibrium in epidemic models with subpopulations*, Math. Biosci. 75(2) (1985), 205–227.

[15] Keshri, A.K., Mishra, B.K. and Rukhaiyar, B.P., *When rumours create chaos in e-commerce*, Chaos Solitons Fractals. 131 (2020) 109497. doi:10.1016/j.chaos.2019.109497.

[16] Khaloufi, I., Benfatah, Y., Laarabi, H. and Rachik, M., *A scenario to fight monkeypox using a mathematical model*, Commun. Math. Biol. Neurosci. 2022 (2022) 99.

[17] Kouidere, A., El Bhih, A., Minifi, I., Balatif, O. and Adnaoui, K., *Optimal control problem for mathematical modeling of Zika virus transmission using fractional order derivatives*, Front. Appl. Math. Stat. 10 (2024), 1376507.

[18] LaSalle, J.P., *The Stability of Dynamical Systems*, Regional Conference Series in Applied Mathematics, Vol. 25, SIAM, Philadelphia, PA, USA, 1976.

[19] Li, H.J., Bu, Z., Wang, Z. and Cao, J., *Dynamical clustering in electronic commerce systems via optimization and leadership expansion*, IEEE Trans. Ind. Inform. 16 (8) (2019) 5327–5334. doi:10.1109/TII.2019.2960835.

[20] Lukes, D.L., *Differential Equations: Classical to Controlled*, 1982.

[21] Lv, R., Li, H., Sun, Q. and Li, B., *Stability analysis and optimal control of a time-delayed panic-spreading model*, Front. Phys. 10 (2022), 1002512.

[22] Makarov, E. and Spitters, B., *The Picard algorithm for ordinary differential equations in Coq*, in *International Conference on Interactive Theorem Proving*, 463–468, Springer, Berlin, Heidelberg, 2013.

[23] Pontryagin, L.S., *The mathematical theory of optimal processes and differential games*, Tr. Mat. Inst. Steklova 169 (1985), 119–158.

[24] Souda, A. and Baiata, M., *Conception et développement d'un module Odoo pour le E-commerce*, Doctoral dissertation, Directeur: M. Fouad Maliki, 2023.

[25] Syniavska, O., *Applied aspects of e-commerce system modeling*, Period. Eng. Nat. Sci. 7(2) (2019), 587–596.

[26] van den Driessche, P. and Watmough, J., *Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission*, Math. Biosci. 180 (2002), 29–48.

Iranian Journal of Numerical Analysis and Optimization
Vol. 17, No. 1, 2027, pp 570–602
https://doi.org/10.22067/ijnao.2026.100432.1943
https://ijnao.um.ac.ir/

CC BY NC

Research Article

Global dynamics of a fractional-order Hantavirus model with awareness-driven protective behavior

Z. Yaagoub*,

Abstract

In this paper, a new fractional-order model of Hantavirus transmission between humans and rodents is proposed, incorporating the effect of media awareness. The model is formulated with the Caputo fractional derivative, which captures the memory effects inherent in epidemic dynamics. The existence, uniqueness, positivity, and boundedness of the solutions are established. The basic reproduction numbers of humans and rodents are derived using the next-generation matrix approach, and the disease-free and endemic equilibrium points are computed. Their global asymptotic stability is established using Lyapunov theory and LaSalle's invariance principle for fractional-order systems. The sensitivity analysis shows that the transmission and recruitment rates have positive effects on the basic reproduction numbers, whereas the recovery and natural mortality rates have negative effects. The numerical results support the theoretical findings on the global stability of the equilibria and show that increasing the effectiveness of media awareness reduces the infection burden. Moreover, the simulations indicate that decreasing the fractional-order parameter leads to faster convergence toward the equilibrium state. However, variations in the fractional-order parameter do not affect the global stability of the system, while increased media awareness contributes to a reduction in Hantavirus infections.

AMS subject classifications (2020): Primary 34A08, 92D30; Secondary 34D20, 37N25, 26A33.

*Corresponding author

Received 12 August 2026; revised 4 October 2026; accepted 9 October 2026

Zakaria Yaagoub
National School of Public Health, Ministry of Health and Social Protection, Rabat, Morocco. e-mail: zakaria.yaag@gmail.com
Department of Computer Science, Faculty of Sciences, Mohammed V University in Rabat, Morocco.

How to cite this article
Yaagoub, Z., Global dynamics of a fractional-order Hantavirus model with awareness-driven protective behavior. *Iran. J. Numer. Anal. Optim.*, 2027; 17(1): 570-602. https://doi.org/10.22067/ijnao.2026.100432.1943

Keywords: Hantavirus; Awareness effect; Fractional model; Global stability; Numerical simulations; Sensitivity analysis.

A Introduction

Hantavirus infection is an emerging zoonotic disease primarily associated with rodent reservoirs and poses a serious public health threat in many parts of the world [17]. Hantavirus is maintained in nature by small mammal reservoir hosts and is transmitted to humans primarily through inhalation of aerosolized rodent urine, saliva, and feces [20]. The transmission dynamics of Hantavirus are influenced by the ecology and distribution of reservoir rodents, as well as by human exposure and preventive behaviors, making disease control challenging [5]. Although much has been learned about the epidemiology and pathogenesis of Hantaviruses, no specific treatment or vaccine is yet available [9]. For this reason, reducing human exposure to infected rodents remains the main means of preventing the disease. Thus, a better understanding of the dynamics of Hantavirus transmission will help public health officials enhance their surveillance systems and develop better prevention and control strategies.

Mathematical modeling has become an essential tool for understanding the dynamics of disease transmission and for supporting public health policy [4, 18]. Mathematical epidemic models integrate biological and epidemiological information to analyze infectious disease transmission. Understanding how infectious diseases spread can help predict disease outbreaks and assess the effectiveness of preventive measures [19]. Over the years, a variety of compartmental models have been developed to study Hantavirus epidemiology in humans and rodents. Such work has improved our understanding of host-reservoir interactions, transmission processes, and the effectiveness of different control strategies on disease persistence [21, 24, 27]. Nevertheless, the complex nature of Hantavirus transmission highlights the need for more refined mathematical models that integrate additional biological mechanisms and behavioral factors. Several mathematical models have been proposed to analyze the transmission dynamics of Hantavirus under different epidemiological assumptions. Early studies used deterministic compartmental models to study the relationship between human and rodent populations and to analyze the impact of demographic and epidemiological factors on disease transmission [16, 26]. Building on this approach, subsequent studies incorporated additional mechanisms such as nonlinear incidence rates, ecological transmission, seasonal patterns, spatial heterogeneity, and stochastic effects [29, 3, 22]. Despite these advances, most current models are still based on classical integer-order differential equations. Such models assume that the future evolution of the epidemic depends only on its current state. Although this assumption is reasonable, it may not fully capture the

memory and hereditary effects present in most biological and epidemiological systems. Davarian et al. [10] investigated a multigroup COVID-19 epidemic model in Iran and analyzed the reproduction number and the stability of its equilibrium points. They also considered optimal control strategies aimed at reducing symptomatic and asymptomatic infections.

To overcome the limitations of traditional integer-order models, fractional-order differential equations have attracted increasing attention in epidemiological modeling over the last two decades. Unlike ordinary derivatives, fractional derivatives are nonlocal and can capture the memory and hereditary effects characteristic of biological and epidemiological systems [15]. This property enables fractional-order models to describe infectious disease transmission more realistically, since the current state of an epidemic depends on its past states and on previous interactions [6]. Accordingly, fractional-order epidemic models have been used to study diseases such as COVID-19, influenza, dengue, tuberculosis, and hepatitis. Several studies have shown that these models can describe epidemiological data and make predictions more accurately than classical integer-order models [1, 7, 31, 2, 8, 33].

Compared with the extensive literature on classical models of Hantavirus transmission, studies based on fractional calculus remain scarce. Nevertheless, the ability of fractional calculus to account for memory and hereditary effects may provide a more realistic picture of Hantavirus dynamics. In this context, Rida et al. [25] developed a fractional-order sensitivity index (SI) model of Hantavirus infection in mice to investigate the effects of environmental conditions and habitat variability on disease dynamics. Using the generalized Euler method, they obtained numerical solutions and showed that changes in environmental conditions, represented by the carrying capacity, can substantially influence the persistence and dynamics of the infected mouse population. Yakit Ongun and Arslan [34] developed explicit and implicit numerical schemes for solving fractional-order Hantavirus models and examined the qualitative properties of their solutions. Moustafa et al. [23] considered a fractional-order Hantavirus infection model with logistic growth, describing the transmission dynamics in a population of susceptible and infected mice. They established the existence, uniqueness, nonnegativity, and boundedness of the solutions. Furthermore, they investigated the local and global asymptotic stability of the equilibrium points and derived the basic reproduction number. The effects of the basic reproduction number and the carrying capacity on the stability of the fractional-order system were also examined through theoretical and numerical analyses. Although these studies demonstrated the advantages of fractional-order models over their integer-order counterparts, they mainly focus on the mathematical and numerical aspects of disease dynamics. In particular, the influence of media-driven public awareness, which can induce protective behavioral responses and thereby reduce human exposure, has not yet been investigated within a fractional-order framework for Hantavirus modeling. The aim of the present study is to bridge this gap by developing a fractional-order model

that incorporates media awareness, providing a more realistic description of Hantavirus transmission.

In light of the above observations, this study presents a new fractional-order Hantavirus transmission model that includes the role of media awareness in influencing disease transmission. The proposed model is formulated using the Caputo fractional derivative to incorporate the memory characteristics of the transmission process. We first prove the positivity, existence, uniqueness, and boundedness of the solutions of the model. Then, we identify the equilibrium points, derive the basic reproduction number, and study the global stability of the equilibria. Afterward, numerical simulations are performed to verify the theoretical results and to investigate the effects of the human transmission rate, the rodent transmission rate, media awareness, and the fractional-order parameter on the spread of Hantavirus. A sensitivity analysis is also conducted to determine the parameters that most influence the disease transmission dynamics. The remainder of this paper is organized as follows: Section B presents the formulation of the fractional-order Hantavirus model. Section C provides proofs of the positivity, existence, uniqueness, and boundedness of the model solutions. Section D discusses the existence of the equilibria and the computation of the basic reproduction number, and derives global stability results for the equilibria. Section E presents the numerical simulation results. It also discusses how the transmission parameters, media awareness, and the fractional-order parameter affect the dynamics of the model. Section F is devoted to the sensitivity analysis. Finally, Section G concludes the paper with a brief summary and future directions.

B Model formulation

Motivated by the need for a better understanding of the role of public awareness in Hantavirus transmission, we propose a new fractional-order epidemic model in the Caputo sense. The model incorporates the role of media through a compartment of protected humans and assumes that greater public awareness reduces the effectiveness of disease transmission, since the public is encouraged to adopt preventive measures. The total human population is divided into four compartments: Susceptible S_h, protected P_h, infectious I_h, and recovered R_h, such that

$$N_h(t) = S_h(t) + P_h(t) + I_h(t) + R_h(t). \tag{1}$$

The rodent population consists of susceptible S_r and infected I_r rodents, satisfying

$$N_r(t) = S_r(t) + I_r(t). \tag{2}$$

Humans acquire infection through effective contact with infected rodents or infectious humans. The parameter η represents the relative contribution of human-to-human transmission. The level of media awareness, denoted by $M(t)$, is generated by infectious individuals. As awareness increases, susceptible individuals adopt precautions and move to the protected class at the rate $\tau M(t)$. Moreover, media awareness further reduces the force of infection through the factor $\frac{1}{1+\kappa M}$. Protected individuals gradually lose their protection and return to the susceptible class at the rate ξ, while recovered individuals lose their temporary immunity at the rate ω. Under these assumptions, the dynamics of Hantavirus transmission are governed by the following system of Caputo fractional differential equations:

$$\begin{cases} D^{\alpha}S_h = \Lambda_h + \omega R_h - \tau M S_h - \dfrac{\beta_h S_h(I_r + \eta I_h)}{1+\kappa M} - \mu_h S_h + \xi P_h, \\ D^{\alpha}P_h = \tau M S_h - (\xi + \mu_h)P_h, \\ D^{\alpha}I_h = \dfrac{\beta_h S_h(I_r + \eta I_h)}{1+\kappa M} - (\gamma_h + \mu_h)I_h, \\ D^{\alpha}R_h = \gamma_h I_h - (\omega + \mu_h)R_h, \\ D^{\alpha}S_r = \Lambda_r - \beta_r S_r I_r - \mu_r S_r, \\ D^{\alpha}I_r = \beta_r S_r I_r - \mu_r I_r, \\ D^{\alpha}M = aI_h - bM. \end{cases} \tag{3}$$

The corresponding initial conditions are $S_h(0) > 0$, $P_h(0) \geq 0$, $I_h(0) \geq 0$, $R_h(0) \geq 0$, $S_r(0) > 0$, $I_r(0) \geq 0$, and $M(0) \geq 0$. The Caputo fractional derivative of order $\alpha \in (0, 1]$ is defined by

$$D^{\alpha}f(t) = \frac{1}{\Gamma(1-\alpha)} \int_0^t \frac{f'(s)}{(t-s)^{\alpha}}\, ds,$$

where $\Gamma(\cdot)$ denotes the Gamma function. The biological meanings of the state variables and model parameters are summarized in Tables 1 and 2. The schematic diagram of the proposed model is shown in Figure 1.

Table 1: Description of the state variables.

Variable	**Description**
$S_h(t)$	Susceptible human population.
$P_h(t)$	Protected human population due to media awareness.
$I_h(t)$	Infectious human population.
$R_h(t)$	Recovered human population with temporary immunity.
$S_r(t)$	Susceptible rodent population.
$I_r(t)$	Infected rodent population.
$M(t)$	Media awareness level.

Table 2: Description of the model parameters.

Parameter	**Description**	**Unit**
Λ_h	Recruitment rate of the human population.	Individuals/day
Λ_r	Recruitment rate of the rodent population.	Rodents/day
μ_h	Natural death rate of humans.	Day^{-1}
μ_r	Natural death rate of rodents.	Day^{-1}
β_h	Effective transmission rate from infected rodents and infectious humans to susceptible humans.	Day^{-1}
β_r	Transmission rate among rodents.	Day^{-1}
η	Modification parameter representing the relative contribution of human-to-human transmission.	–
τ	Protection rate induced by media awareness.	Day^{-1}
κ	Media effectiveness parameter reducing disease transmission.	–
ξ	Rate at which protected individuals return to the susceptible class.	Day^{-1}
γ_h	Recovery rate of infectious humans.	Day^{-1}
ω	Loss of immunity rate of recovered humans.	Day^{-1}
a	Media awareness generation rate.	Day^{-1}
b	Natural decay rate of media awareness.	Day^{-1}
α	Fractional-order parameter, $0 < \alpha \leq 1$.	–

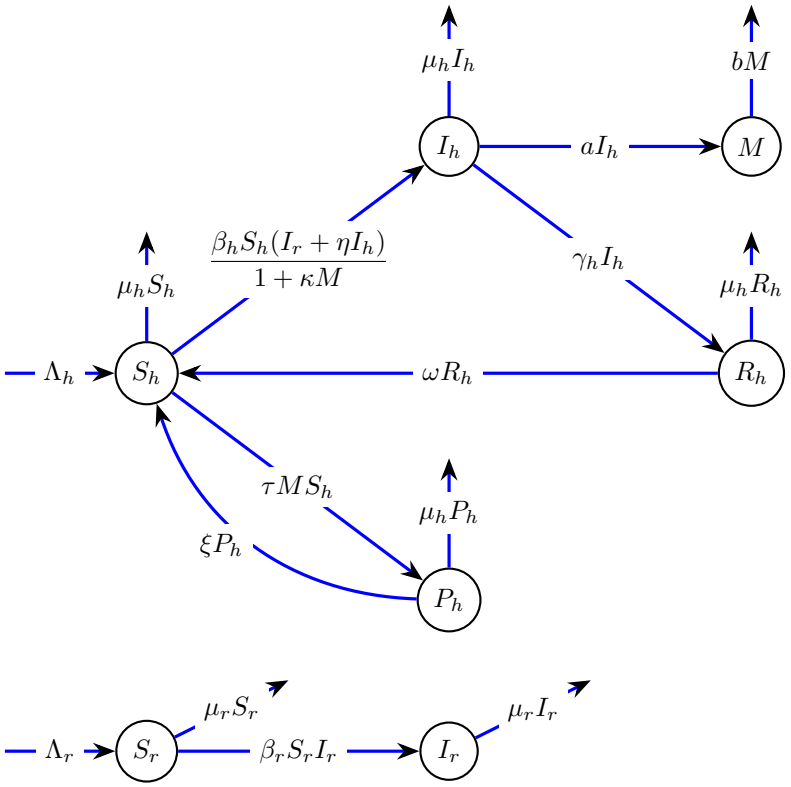


Figure 1: Flow diagram of the fractional-order Hantavirus transmission model with awareness dynamics.

C Mathematical well-posedness of the model

In this section, we verify the well-posedness of system (3). Specifically, we prove the existence and uniqueness of a positive solution of the fractional-order model and show that the state variables remain bounded for all $t > 0$.

Proposition 1. For any positive initial conditions, the variables of system (3), namely $S_h(t)$, $P_h(t)$, $I_h(t)$, $R_h(t)$, $S_r(t)$, $I_r(t)$, and $M(t)$, remain positive for all $t > 0$.

Proof. From the different components of system (3), we have

$$\begin{cases} D^{\alpha}S_h\mid_{S_h=0} = \Lambda_h + \omega R_h + \xi P_h \geq 0, \\ D^{\alpha}P_h\mid_{P_h=0} = \tau M S_h \geq 0, \\ D^{\alpha}I_h\mid_{I_h=0} \geq 0, \\ D^{\alpha}R_h\mid_{R_h=0} = \gamma_h I_h \geq 0, \\ D^{\alpha}S_r\mid_{S_r=0} = \Lambda_r \geq 0, \\ D^{\alpha}I_r\mid_{I_r=0} = 0, \\ D^{\alpha}M\mid_{M=0} = aI_h \geq 0. \end{cases} \tag{4}$$

By applying [1, Lemmas 2 and 3], it follows that every solution of system (3) with positive initial conditions remains positive for all $t > 0$. Hence, the proposed model preserves the positivity of its solutions, which ensures that it is both biologically meaningful and mathematically well-defined. □

Next, we investigate the boundedness of the state variables of system (3). The following proposition establishes that all solutions of the system remain bounded.

Proposition 2. The biologically feasible region

$$H = \Big\{ (S_h, P_h, I_h, R_h, S_r, I_r, M) \in \mathbb{R}_+^7 : \\ S_h + P_h + I_h + R_h \leq \frac{\Lambda_h}{\mu_h}, \quad S_r + I_r \leq \frac{\Lambda_r}{\mu_r}, \quad M \leq \frac{a\Lambda_h}{b\mu_h} \Big\}$$

is positively invariant.

Proof. Let

$$N_h(t) = S_h(t) + P_h(t) + I_h(t) + R_h(t).$$

By summing the first four equations of system (3), we obtain

$$\begin{aligned} D^{\alpha}N_h(t) &= D^{\alpha}S_h + D^{\alpha}P_h + D^{\alpha}I_h + D^{\alpha}R_h \\ &\leq \Lambda_h - \mu_h S_h - \mu_h P_h - \mu_h I_h - \mu_h R_h \\ &\leq \Lambda_h - \mu_h\big(S_h + P_h + I_h + R_h\big). \end{aligned}$$

Thus,

$$D^{\alpha}N_h(t) \leq \Lambda_h - \mu_h N_h(t).$$

Therefore, we have

$$\limsup_{t\to+\infty} N_h(t) \leq \frac{\Lambda_h}{\mu_h}.$$

Similarly, let

$$N_r(t) = S_r(t) + I_r(t).$$

By summing the rodent equations of system (3), we have

$$\begin{aligned} D^\alpha N_r(t) &= D^\alpha S_r + D^\alpha I_r \\ &\leq \Lambda_r - \mu_r S_r - \mu_r I_r. \end{aligned}$$

Hence,

$$D^\alpha N_r(t) \leq \Lambda_r - \mu_r N_r(t),$$

which implies

$$\limsup_{t\to+\infty} N_r(t) \leq \frac{\Lambda_r}{\mu_r}.$$

Now, from the last equation of system (3), we obtain

$$D^\alpha M = aI_h - bM.$$

Since

$$I_h \leq N_h(t) \leq \frac{\Lambda_h}{\mu_h},$$

we have

$$D^\alpha M \leq \frac{a\Lambda_h}{\mu_h} - bM.$$

Therefore,

$$\limsup_{t\to+\infty} M(t) \leq \frac{a\Lambda_h}{b\mu_h}.$$

Consequently, the region H is positively invariant. □

Theorem 1. For every initial condition in H, system (3) has a unique solution.

Proof. First, system (3) can be rewritten as

$$D^{\alpha}X(t) = F(t, X(t)),$$

where

$$X(t) = \left(S_h(t), P_h(t), I_h(t), R_h(t), S_r(t), I_r(t), M(t)\right),$$

and

$$F(t, X(t)) = \begin{pmatrix} \Lambda_h + \omega R_h - \tau M S_h - \dfrac{\beta_h S_h(I_r + \eta I_h)}{1 + \kappa M} - \mu_h S_h + \xi P_h \\ \tau M S_h - (\xi + \mu_h) P_h \\ \dfrac{\beta_h S_h(I_r + \eta I_h)}{1 + \kappa M} - (\gamma_h + \mu_h) I_h \\ \gamma_h I_h - (\omega + \mu_h) R_h \\ \Lambda_r - \beta_r S_r I_r - \mu_r S_r \\ \beta_r S_r I_r - \mu_r I_r \\ a I_h - b M \end{pmatrix}.$$

The function $F(t, X(t))$ satisfies the first condition of [32, Proposition 4]. To verify the second condition of this proposition, we use the boundedness of the solutions established above, which gives

$$\begin{aligned} \|F(t, X(t))\|_1 \leq & \Lambda_h + \Lambda_r + \left(\frac{2a\tau\Lambda_h}{b\mu_h} + \frac{2b\beta_h\mu_h(\Lambda_r\mu_h + \eta\Lambda_h\mu_r)}{b\mu_h + \kappa a\Lambda_h} + \mu_h \right) \|S_h(t)\|_1 \\ & + (2\xi + \mu_h) \|P_h(t)\|_1 + (2\gamma_h + a + \mu_h) \|I_h(t)\|_1 \\ & + (2\omega + \mu_h) \|R_h(t)\|_1 + \left(\frac{2\beta_r\Lambda_r}{\mu_r} + \mu_r \right) \|S_r(t)\|_1 + \mu_r \|I_r(t)\|_1 \\ & + b\|M\|_1 \leq \Lambda_T + \Theta \|X(t)\|_1, \end{aligned}$$

with

$$\Lambda_T = \Lambda_h + \Lambda_r$$

and

$$\Theta = \max\left\{\frac{2a\tau\Lambda_h}{b\mu_h} + \frac{2b\beta_h\mu_h(\Lambda_r\mu_h + \eta\Lambda_h\mu_r)}{b\mu_h + \kappa a\Lambda_h} + \mu_h,\ 2\xi + \mu_h,\right.$$
$$\left. 2\gamma_h + a + \mu_h,\ 2\omega + \mu_h,\ \frac{2\beta_r\Lambda_r}{\mu_r} + \mu_r,\ \mu_r,\ b\right\}.$$

Hence, $F(t, X(t))$ satisfies both conditions of [32, Proposition 4]. Consequently, the fractional system (3) admits a unique solution in the region H. This completes the proof. □

D Global stability analysis of the equilibria

This section analyzes the existence of the equilibrium points of the model, namely the disease-free equilibrium E_0, the human-only equilibrium E_h, the rodent-only equilibrium E_r, and the endemic equilibrium E^*. In addition, the basic reproduction number R_0 is computed using the next-generation matrix technique, and the conditions for the existence of each equilibrium point are derived. Moreover, using Lyapunov functions together with LaSalle's invariance principle, we prove the global stability of these equilibrium points. We begin with the derivation of the basic reproduction number associated with system (3).

D.1 Computation of the basic reproduction number

The basic reproduction number, denoted by R_0, is defined as the average number of secondary infections produced by a single infected individual introduced into a fully susceptible population [28]. We use the next-generation matrix method as in [30] to calculate R_0, which is given by

$$R_0 = \rho(FV^{-1}),$$

where F is the nonnegative matrix of new infection terms, V is the matrix of the remaining transition terms in system (3), and ρ denotes the spectral radius of the matrix FV^{-1}. Let

$$F = \begin{pmatrix} \frac{\beta_h\eta\Lambda_h}{\mu_h} & \frac{\beta_h\Lambda_h}{\mu_h} \\ 0 & \frac{\beta_r\Lambda_r}{\mu_r} \end{pmatrix}, \quad V = \begin{pmatrix} \gamma_h + \mu_h & 0 \\ 0 & \mu_r \end{pmatrix}.$$

Therefore, by computing the matrix FV^{-1}, we obtain

$$R_0 = \max\left\{R_0^h, R_0^r\right\},$$

where

$$\begin{cases} R_0^h = \dfrac{\beta_h \eta \Lambda_h}{\mu_h(\gamma_h + \mu_h)}, \\ R_0^r = \dfrac{\beta_r \Lambda_r}{\mu_r^2}. \end{cases}$$

Here,

- R_0^h represents the human-related basic reproduction number associated with the human-to-human transmission pathway;
- R_0^r denotes the rodent-related basic reproduction number associated with the transmission dynamics among rodents.

Consequently, the overall basic reproduction number R_0 is determined by the dominant transmission mechanism between the human and rodent populations.

D.2 Determination of steady states

Model (3) admits four equilibrium points: The disease-free equilibrium E_0, the human-only equilibrium E_h, the rodent-only equilibrium E_r, and the endemic equilibrium E^*. These equilibrium states are given as follows:

1. **Disease-free equilibrium point.** The disease-free equilibrium point is given by

$$E_0 = \left(S_h^0,\, P_h^0,\, I_h^0,\, R_h^0,\, S_r^0,\, I_r^0,\, M^0\right),$$

where

$$S_h^0 = \frac{\Lambda_h}{\mu_h}, \quad S_r^0 = \frac{\Lambda_r}{\mu_r}, \quad P_h^0 = I_h^0 = R_h^0 = I_r^0 = M^0 = 0.$$

The disease-free equilibrium point E_0 always exists, since all its components are nonnegative.

2. **Human-only equilibrium point.** The human-only equilibrium point, denoted by E_h, represents a state in which the infection persists exclusively within the human population, while it is absent from the rodent population. It is given by

$$E_h = (S_h^h, P_h^h, I_h^h, R_h^h, S_r^h, 0, M^h),$$

where

$$S_r^h = \frac{\Lambda_r}{\mu_r}, \quad M^h = \frac{a}{b} I_h^h, \quad R_h^h = \frac{\gamma_h}{\omega + \mu_h} I_h^h, \quad P_h^h = \frac{\tau M^h S_h^h}{\xi + \mu_h},$$

and

$$S_h^h = \frac{(\gamma_h + \mu_h)(1 + \kappa M^h)}{\beta_h \eta}.$$

Furthermore, I_h^h is obtained as a positive solution of the corresponding equilibrium equation. Since $I_h^h > 0$, the infected human equilibrium equation is given by

$$\frac{\beta_h \eta S_h^h}{1 + \kappa M^h} = \gamma_h + \mu_h.$$

Therefore, a positive human-only equilibrium exists whenever

$$R_0^h = \frac{\beta_h \eta \Lambda_h}{\mu_h(\gamma_h + \mu_h)} > 1.$$

3. **Rodent-only equilibrium point.** The rodent-only equilibrium point, denoted by E_r, corresponds to the persistence of the infection in the rodent population while the human population remains disease-free. It is given by

$$E_r = (S_h^r, 0, 0, 0, S_r^r, I_r^r, 0),$$

where

$$S_h^r = \frac{\Lambda_h}{\mu_h}, \quad S_r^r = \frac{\mu_r}{\beta_r}, \quad I_r^r = \frac{\Lambda_r \beta_r - \mu_r^2}{\beta_r \mu_r}.$$

Furthermore, I_r^r is obtained as a positive solution of the corresponding equilibrium equation. Since $I_r^r > 0$, the infected rodent equilibrium equation is given by

$$\beta_r S_r^r = \mu_r.$$

Substituting $S_r^r = \frac{\mu_r}{\beta_r}$ into the susceptible rodent equilibrium equation gives

$$I_r^r = \frac{\Lambda_r \beta_r - \mu_r^2}{\beta_r \mu_r}.$$

Therefore, a positive rodent-only equilibrium exists whenever

$$\beta_r \Lambda_r - \mu_r^2 > 0.$$

Hence, the rodent-only equilibrium point E_r exists if and only if

$$R_0^r = \frac{\beta_r \Lambda_r}{\mu_r^2} > 1.$$

4. **Endemic equilibrium point.** The endemic equilibrium point, denoted by E^*, represents a steady state in which the infection persists in both the human and rodent populations. It is given by

$$E^* = (S_h^*, P_h^*, I_h^*, R_h^*, S_r^*, I_r^*, M^*),$$

where

$$S_r^* = \frac{\mu_r}{\beta_r}, \quad I_r^* = \frac{\Lambda_r \beta_r - \mu_r^2}{\beta_r \mu_r}, \quad M^* = \frac{a}{b} I_h^*,$$
$$R_h^* = \frac{\gamma_h}{\omega + \mu_h} I_h^*, \quad P_h^* = \frac{\tau M^* S_h^*}{\xi + \mu_h},$$

and

$$S_h^* = \frac{(\gamma_h + \mu_h) I_h^* (1 + \kappa M^*)}{\beta_h (I_r^* + \eta I_h^*)}.$$

The quantity I_h^* is determined as a positive solution of the corresponding equilibrium equation. For the endemic equilibrium to be biologically meaningful, all state variables must be positive. In particular, since $I_r^* > 0$, we require

$$\Lambda_r \beta_r - \mu_r^2 > 0,$$

which can be rewritten as

$$R_0^r = \frac{\beta_r \Lambda_r}{\mu_r^2} > 1.$$

Similarly, persistence of the disease in the human population requires a positive infected class, $I_h^* > 0$. This condition is satisfied whenever

$$R_0^h = \frac{\beta_h \eta \Lambda_h}{\mu_h (\gamma_h + \mu_h)} > 1.$$

Therefore, the endemic equilibrium E^* exists whenever both the rodent and human basic reproduction numbers exceed unity, that is,

$$R_0^h > 1 \qquad \text{and} \qquad R_0^r > 1.$$

The existence conditions of the equilibrium points of system (3) are summarized in Table 3.

Table 3: Summary of the existence conditions for the equilibrium points of system (3).

Equilibrium Point	**Description**	**Existence Condition**
E_0	Disease-free equilibrium	Always exists
E_h	Human-only equilibrium	$R_0^h > 1$
E_r	Rodent-only equilibrium	$R_0^r > 1$
E^*	Endemic equilibrium	$R_0^h > 1$ and $R_0^r > 1$

Next, we investigate the global stability properties of the equilibrium points of system (3).

D.3 Global stability

This subsection investigates the global stability of the equilibrium points of system (3) using Lyapunov's direct method and LaSalle's invariance principle [12]. Although a human-only equilibrium E_h may exist when $R_0^h > 1$, it is of limited relevance to Hantavirus transmission, because rodents are the natural reservoir of the virus and the long-term persistence of the disease depends on the survival and proliferation of infected rodent populations. Thus, the equilibrium point E_h, which excludes the persistence of infection in rodents, does not accurately reflect the key dynamics of the coupled human-rodent transmission system. Similarly, the rodent-only equilibrium E_r, in which rodents remain infected while humans are free of infection, is not the primary focus of this study. This equilibrium may be biologically feasible, but it provides limited insight into the public health consequences of Hantavirus infection. Since our main aim is to determine the mechanisms governing disease spread and persistence in both host populations, the global stability analysis is restricted to the disease-free equilibrium E_0 and the endemic equilibrium E^*. These equilibria represent the most epidemiologically significant states of the system and are crucial in determining the threshold dynamics and the long-term behavior of the Hantavirus transmission system.

Theorem 2. If $R_0 = \max\{R_0^h, R_0^r\} \leq 1$, then the disease-free equilibrium point E_0 of system (3) is globally asymptotically stable in the feasible region H.

Proof. Consider the following function:

$$V_0 = \frac{I_h}{\gamma_h + \mu_h} + \frac{I_r}{\mu_r}.$$

Clearly, V_0 is continuous and nonnegative throughout the feasible region H. Since $I_h \geq 0$ and $I_r \geq 0$, it follows directly that $V_0 \geq 0$. Moreover, $V_0 = 0$ if and only if $I_h = I_r = 0$, which implies that V_0 is positive definite with respect to the disease-free equilibrium E_0. Consequently, V_0 is a suitable Lyapunov function candidate for analyzing the global stability of E_0. By applying the properties of the Caputo fractional derivative, we obtain

$$D^\alpha V_0 \leq \frac{1}{\gamma_h + \mu_h} D^\alpha I_h + \frac{1}{\mu_r} D^\alpha I_r.$$

Substituting the equations of system (3), we get

$$D^\alpha V_0 \leq \frac{1}{\gamma_h + \mu_h} \left(\frac{\beta_h S_h (I_r + \eta I_h)}{1 + \kappa M} - (\gamma_h + \mu_h) I_h \right) + \frac{1}{\mu_r} \left(\beta_r S_r I_r - \mu_r I_r \right).$$

Since

$$\frac{1}{1 + \kappa M} \leq 1, \qquad S_h \leq \frac{\Lambda_h}{\mu_h}, \qquad S_r \leq \frac{\Lambda_r}{\mu_r},$$

it follows that

$$D^\alpha V_0 \leq \frac{\beta_h \Lambda_h}{\mu_h (\gamma_h + \mu_h)} I_r + \left(\frac{\beta_h \eta \Lambda_h}{\mu_h (\gamma_h + \mu_h)} - 1 \right) I_h + \left(\frac{\beta_r \Lambda_r}{\mu_r^2} - 1 \right) I_r.$$

Using the definitions

$$R_0^h = \frac{\beta_h \eta \Lambda_h}{\mu_h (\gamma_h + \mu_h)} \quad \text{and} \quad R_0^r = \frac{\beta_r \Lambda_r}{\mu_r^2},$$

we obtain

$$D^\alpha V_0 \leq (R_0^h - 1) I_h + (R_0^r - 1) I_r.$$

Since $R_0^h \leq 1$ and $R_0^r \leq 1$, it follows that $D^\alpha V_0 \leq 0$. Moreover, $D^\alpha V_0 = 0$ if and only if $I_h = 0$ and $I_r = 0$. From the system equations, when $I_h = I_r = 0$, we necessarily obtain $M = 0$, $P_h = 0$, and $R_h = 0$, while $S_h = \frac{\Lambda_h}{\mu_h}$ and $S_r = \frac{\Lambda_r}{\mu_r}$. Therefore, the largest invariant set contained

in $\{(I_h, I_r) : D^\alpha V_0 = 0\}$ is the singleton $\{E_0\}$. Consequently, by LaSalle's invariance principle for fractional-order systems, the disease-free equilibrium E_0 is globally asymptotically stable in H. □

Theorem 3. If $R_0^h > 1$ and $R_0^r > 1$, then the endemic equilibrium point E^* is globally asymptotically stable in the feasible region H.

Proof. Consider the following function:

$$\begin{aligned} V^* = {} & S_h^* \left(\frac{S_h}{S_h^*} - \ln \frac{S_h}{S_h^*} - 1 \right) + P_h^* \left(\frac{P_h}{P_h^*} - \ln \frac{P_h}{P_h^*} - 1 \right) + I_h^* \left(\frac{I_h}{I_h^*} - \ln \frac{I_h}{I_h^*} - 1 \right) \\ & + R_h^* \left(\frac{R_h}{R_h^*} - \ln \frac{R_h}{R_h^*} - 1 \right) + S_r^* \left(\frac{S_r}{S_r^*} - \ln \frac{S_r}{S_r^*} - 1 \right) \\ & + I_r^* \left(\frac{I_r}{I_r^*} - \ln \frac{I_r}{I_r^*} - 1 \right) + M^* \left(\frac{M}{M^*} - \ln \frac{M}{M^*} - 1 \right). \end{aligned}$$

Clearly, V^* is continuous and nonnegative throughout the feasible region H. Indeed, for every positive variable x, the inequality

$$x - \ln(x) - 1 \geq 0$$

holds, with equality if and only if $x = 1$. Since each component of V^* is of this form, every term is nonnegative, implying that

$$V^* \geq 0 \quad \text{for all } (S_h, P_h, I_h, R_h, S_r, I_r, M) \in H.$$

Moreover, $V^* = 0$ if and only if $S_h = S_h^*$, $P_h = P_h^*$, $I_h = I_h^*$, $R_h = R_h^*$, $S_r = S_r^*$, $I_r = I_r^*$, $M = M^*$, which is equivalent to

$$(S_h, P_h, I_h, R_h, S_r, I_r, M) = E^*.$$

Therefore, V^* vanishes only at the endemic equilibrium E^* and remains strictly positive elsewhere in H. Hence, V^* is positive definite with respect to E^*, which makes it an appropriate Lyapunov function candidate for establishing the global asymptotic stability of the endemic equilibrium point E^*. By [1, Lemma 4], we obtain

$$\begin{aligned} D^\alpha V^* \leq {} & \left(1 - \frac{S_h^*}{S_h}\right) D^\alpha S_h + \left(1 - \frac{P_h^*}{P_h}\right) D^\alpha P_h + \left(1 - \frac{I_h^*}{I_h}\right) D^\alpha I_h \\ & + \left(1 - \frac{R_h^*}{R_h}\right) D^\alpha R_h + \left(1 - \frac{S_r^*}{S_r}\right) D^\alpha S_r + \left(1 - \frac{I_r^*}{I_r}\right) D^\alpha I_r \end{aligned}$$

$$+\left(1-\frac{M^*}{M}\right)D^\alpha M.$$

Substituting the model equations into the above inequality yields

$$\begin{aligned}
D^\alpha V^* \le{}& \left(1-\frac{S_h^*}{S_h}\right)\left(\Lambda_h+\omega R_h-\tau M S_h-\frac{\beta_h S_h(I_r+\eta I_h)}{1+\kappa M}-\mu_h S_h+\xi P_h\right)\\
&+\left(1-\frac{P_h^*}{P_h}\right)(\tau M S_h-(\xi+\mu_h)P_h)\\
&+\left(1-\frac{I_h^*}{I_h}\right)\left(\frac{\beta_h S_h(I_r+\eta I_h)}{1+\kappa M}-(\gamma_h+\mu_h)I_h\right)\\
&+\left(1-\frac{R_h^*}{R_h}\right)(\gamma_h I_h-(\omega+\mu_h)R_h)\\
&+\left(1-\frac{S_r^*}{S_r}\right)(\Lambda_r-\beta_r S_r I_r-\mu_r S_r)\\
&+\left(1-\frac{I_r^*}{I_r}\right)(\beta_r S_r I_r-\mu_r I_r)\\
&+\frac{\gamma_h+\mu_h}{a}\left(1-\frac{M^*}{M}\right)(aI_h-bM).
\end{aligned}$$

Using the equilibrium relations

$$\begin{cases}
\tau M^* S_h^* = (\xi+\mu_h)P_h^*,\\
\dfrac{\beta_h S_h^*(I_r^*+\eta I_h^*)}{1+\kappa M^*} = (\gamma_h+\mu_h)I_h^*,\\
\gamma_h I_h^* = (\omega+\mu_h)R_h^*,\\
\beta_r S_r^* I_r^* = \mu_r I_r^*,\\
aI_h^* = bM^*,
\end{cases} \tag{5}$$

and after straightforward calculations, we obtain

$$D^\alpha V^* \le -F_h^*\Phi_1 - G_h^*\Phi_2 - \gamma_h I_h^*\Phi_3 - F_r^*\Phi_4 - bM^*\Phi_5,$$

where

$$\begin{cases}
F_h^* = \dfrac{\beta_h S_h^*(I_r^*+\eta I_h^*)}{1+\kappa M^*},\\
G_h^* = \tau M^* S_h^*,\\
F_r^* = \beta_r S_r^* I_r^*,
\end{cases} \tag{6}$$

and

$$\begin{cases} \Phi_1 = 3 - \dfrac{S_h^*}{S_h} - \dfrac{S_h(I_r+\eta I_h)(1+\kappa M^*)I_h^*}{S_h^*(I_r^*+\eta I_h^*)(1+\kappa M)I_h} - \dfrac{I_h S_h^*(I_r^*+\eta I_h^*)(1+\kappa M)}{I_h^* S_h(I_r+\eta I_h)(1+\kappa M^*)}, \\ \Phi_2 = 3 - \dfrac{S_h^*}{S_h} - \dfrac{M S_h P_h^*}{M^* S_h^* P_h} - \dfrac{P_h M^* S_h^*}{P_h^* M S_h}, \\ \Phi_3 = 2 - \dfrac{I_h R_h^*}{I_h^* R_h} - \dfrac{R_h I_h^*}{R_h^* I_h}, \\ \Phi_4 = 2 - \dfrac{S_r^*}{S_r} - \dfrac{S_r I_r^*}{S_r^* I_r}, \\ \Phi_5 = 2 - \dfrac{I_h M^*}{I_h^* M} - \dfrac{M I_h^*}{M^* I_h}. \end{cases} \tag{7}$$

By the arithmetic-geometric mean inequality,

$$\Phi_i \geq 0, \qquad i = 1, \ldots, 5.$$

Therefore,

$$D^\alpha V^* \leq 0 \quad \text{for all } (S_h, P_h, I_h, R_h, S_r, I_r, M) \in H.$$

Moreover, $D^\alpha V^* = 0$ if and only if $S_h = S_h^*$, $P_h = P_h^*$, $I_h = I_h^*$, $R_h = R_h^*$, $S_r = S_r^*$, $I_r = I_r^*$, and $M = M^*$. Therefore, every state variable coincides with its corresponding endemic equilibrium value, which implies that the largest invariant set contained in $\{D^\alpha V^* = 0\}$ is the singleton $\{E^*\}$. Consequently, by LaSalle's invariance principle for fractional-order systems, every solution of system (3) with initial conditions in H approaches E^* as t tends to infinity. Hence, the endemic equilibrium point E^* is globally asymptotically stable in H. □

E Numerical simulations

In this section, a series of numerical simulations is carried out to investigate the influence of the fractional order and of the transmission and awareness parameters on the infection dynamics. All simulations were performed using MATLAB. To obtain approximate solutions of the proposed fractional-order model, we employed the numerical scheme introduced in [14]. The discretization formula is given by

$$\begin{aligned} Y(t_i) = Y(t_0) &+ \frac{h^\alpha}{\Gamma(\alpha+2)}\left[(i-1)^{\alpha+1} - (i-\alpha-1)i^\alpha\right] G\big(Y(t_0)\big) \\ &+ \frac{h^\alpha}{\Gamma(\alpha+2)} \sum_{j=1}^{i-1} \left[(i-j+1)^{\alpha+1} - 2(i-j)^{\alpha+1} + (i-j-1)^{\alpha+1}\right] G\big(Y(t_j)\big) \end{aligned}$$

$$+\frac{h^{\alpha}}{\Gamma(\alpha+2)}G\left(Y(t_{i-1})+\frac{h^{\alpha}}{\Gamma(\alpha+1)}F\big(Y(t_{i-1})\big)\right), \tag{8}$$

where the discrete time nodes are defined by

$$t_i = t_0 + ih, \qquad i = 0, 1, \ldots, N-1,$$

with h denoting the time-step size. The initial condition is

$$Y_0 = (50,\ 30,\ 10,\ 5,\ 30,\ 7,\ 15).$$

Unless otherwise specified, the parameter values used in the numerical simulations are listed in Table 4. These values were used to generate all the figures presented in this section.

Table 4: Parameter values used in the numerical simulations.

Parameters	Figure 2	Figure 3	Figure 4	Figure 5	Figure 6
Λ_h	1	1	1	1	1
μ_h	0.05	0.1	0.1	0.1	0.1
β_h	0.0002	0.25	-	0.25	0.25
η	0.1	0.8	0.8	0.8	0.8
κ	0.1	0.5	0.5	0.5	-
τ	0.02	0.002	0.002	0.002	0.002
ξ	0.01	0.05	0.05	0.05	0.05
γ_h	0.3	0.08	0.08	0.08	0.08
ω	0.05	0.05	0.05	0.05	0.05
Λ_r	1	1	1	1	1
β_r	0.0003	0.2	0.2	-	0.2
μ_r	0.1	0.25	0.25	0.25	0.25
a	0.05	0.02	0.02	0.02	0.02
b	0.3	0.20	0.20	0.20	0.20
R_0^h	0.00114	11.11	-	-	-
R_0^r	0.03	3.20	-	-	-
Theorems	Theorem 2	Theorem 3	-	-	-

E.1 Global stability and effect of the fractional order

To illustrate the global stability of the equilibria and the effect of the fractional order on the infection dynamics, we used scheme (8), with the fractional order ranging between 0.55 and 1. Figure 2 illustrates the time evolution of the Hantavirus model variables, namely S_h, P_h, I_h, R_h, S_r, I_r, and M, for different values of the fractional-order parameter α. As shown, all state trajectories converge to the disease-free equilibrium $E_0 = (20, 0, 0, 0, 10, 0, 0)$ for all values of α considered. This behavior indicates the complete elimination of the infection from both the human and rodent populations, with only the susceptible classes remaining positive at equilibrium. These numerical results are fully consistent with Theorem 2, which guarantees the global asymptotic stability of the disease-free equilibrium. Furthermore, the basic reproduction numbers reported in Table 4, $R_0^h = 0.00114 < 1$ and $R_0^r = 0.03 < 1$, satisfy the conditions of Theorem 2.

Consequently, the infection cannot persist in either the human or the rodent population, and the disease-free state is globally attractive. The simulations also show that the fractional-order parameter has a significant impact on the transient dynamics of the system. In particular, lower values of α lead to faster convergence to the equilibrium, whereas higher values lead to slower convergence. This phenomenon is due to the memory effect characteristic of fractional-order derivatives: The higher the fractional order, the stronger the memory of the system and the longer the transient response before the steady state is reached. It should be noted that changing the fractional-order parameter alters neither the location nor the stability of the equilibrium point. In other words, the fractional order only influences the convergence speed of the trajectories, while the long-term qualitative behavior remains unchanged. Figure 3 illustrates the time evolution of the Hantavirus model variables for different values of α. The numerical results show that, irrespective of the chosen fractional order, all solution trajectories converge to the endemic equilibrium point $E^* = (0.820, 0.009, 5.983, 3.191, 1.250, 2.750, 0.873)$. This convergence demonstrates the long-term persistence of the disease in both the human and rodent populations. In particular, the infected human and rodent compartments approach positive equilibrium values, confirming that the infection remains endemic. Furthermore, the protected human compartment P_h stabilizes at a very small positive level, which can be attributed to the relatively low awareness rate τ. These numerical observations are fully consistent with Theorem 3, which establishes the global asymptotic stability of the endemic equilibrium. Moreover, as reported in Table 4, the basic reproduction numbers satisfy $R_0^h = 11.11 > 1$ and $R_0^r = 3.20 > 1$, thereby fulfilling the conditions of Theorem 3.

Consequently, the endemic equilibrium exists and is globally asymptotically stable. The simulations also emphasize the role of the fractional-order parameter in the transient behavior

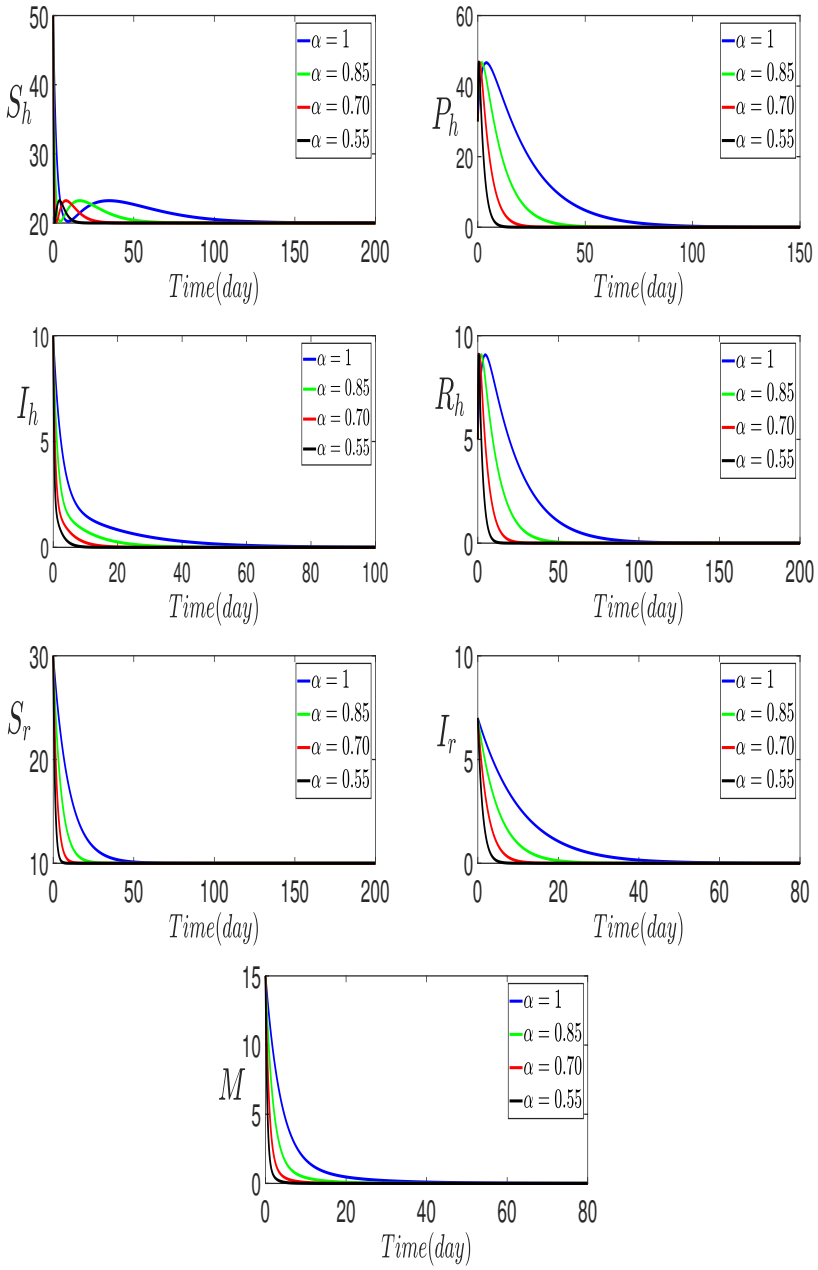


Figure 2: Stability of the disease-free equilibrium point E_0 for different values of the fractional-order parameter α.

of the system. In particular, lower fractional-order values result in faster convergence to the endemic equilibrium, while higher values lead to a longer transient phase. This behavior is typical

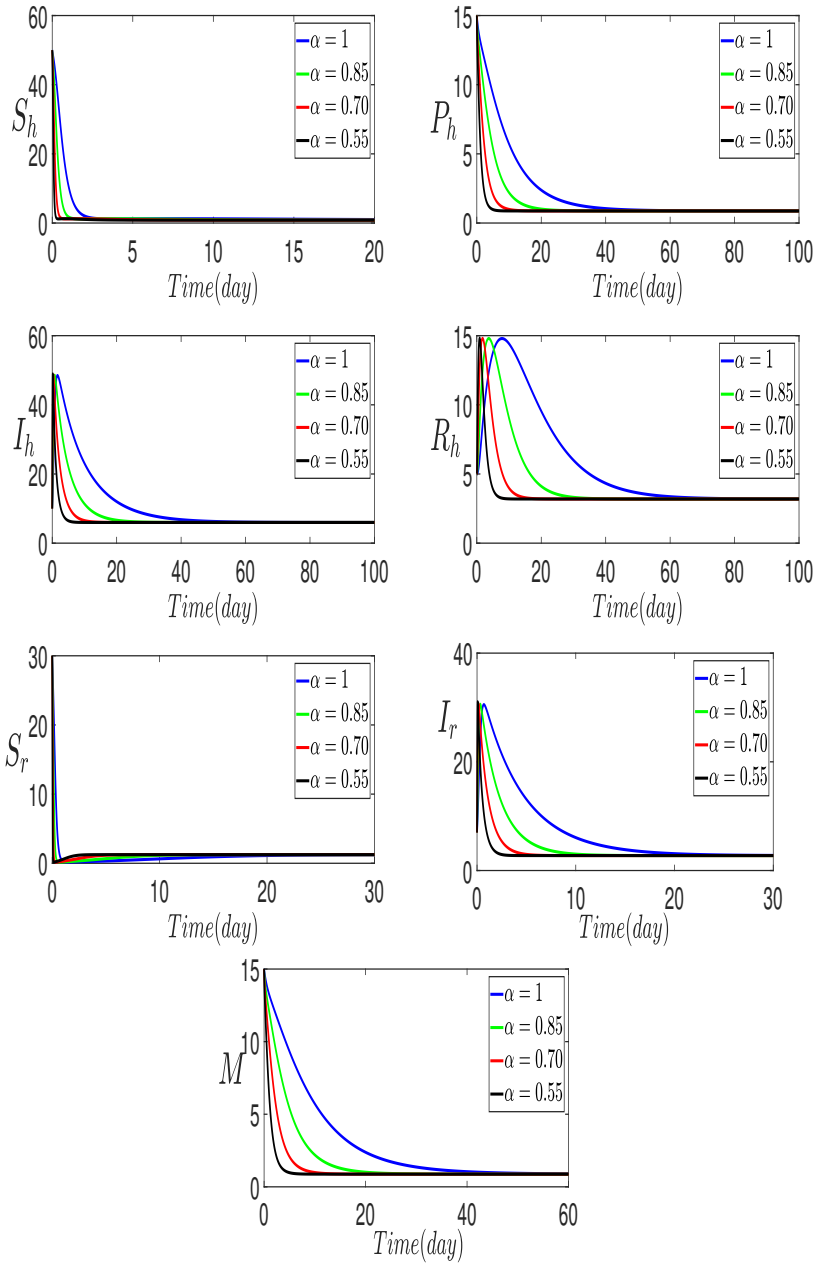


Figure 3: Stability of the endemic equilibrium point E^* for different values of the fractional-order parameter α.

of fractional-order dynamical systems, in which memory effects arise from the Caputo fractional derivative. As the order increases, the memory effect becomes more significant, which translates

into a longer transient period before the system reaches equilibrium. However, changing the value of the fractional-order parameter alters neither the position nor the stability of the endemic equilibrium; it only affects the rate at which the solution trajectories converge to it.

E.2 Effects of the transmission rates

In this subsection, we study how the transmission rates affect the dynamics of the Hantavirus model. The transmission rates play a key role in the spread of the disease among humans and rodents. We analyze the impact of varying these parameters on the evolution of the model variables while keeping the other parameters unchanged.

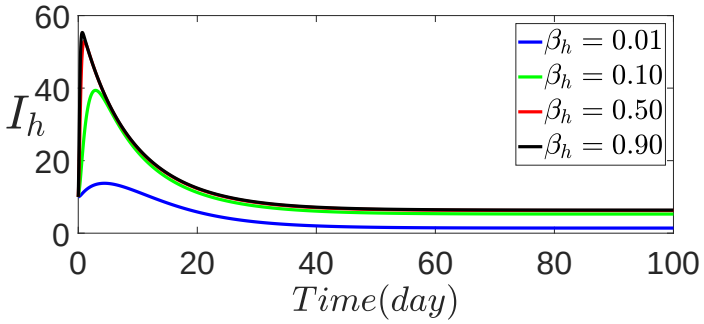


Figure 4: Influence of the human transmission rate β_h on the infected human population I_h.

Figure 4 highlights the impact of the human transmission rate β_h on the infected human population I_h. The results show that, as β_h increases, the number of infected individuals increases throughout the simulation. For small values of β_h, the number of infections decreases quickly, leading to a low endemic level. In contrast, larger values of β_h result in a high infection peak and a high endemic equilibrium. These findings demonstrate that reducing the human transmission rate is an effective means of preventing the spread of Hantavirus and reducing its burden on the human population over time. Figure 5 presents the impact of the rodent transmission rate β_r on the infected rodent population. It can be seen that increasing β_r increases the number of infected rodents throughout the simulation, whereas decreasing this parameter substantially reduces the infection level among rodents. Although all trajectories eventually converge to an equilibrium, the transient dynamics are considerably affected by the value of β_r. For instance, for larger values of β_r, the infection level is higher and persists longer before the steady state is reached. Thus, this figure illustrates the importance of the rodent transmission rate in maintaining and propagating Hantavirus.

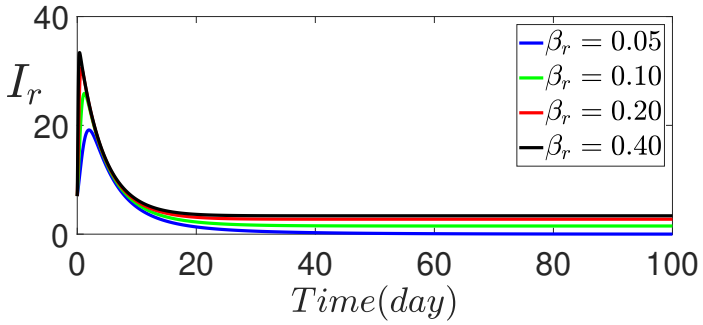


Figure 5: Impact of the rodent transmission rate β_r on the infected rodent population I_r.

E.3 Effect of media awareness

In this subsection, we analyze the effect of media awareness, which encourages individuals to adopt preventive behaviors and thereby reduces the infection rate. By varying the media awareness parameter while keeping all other parameters unchanged, we assess its effect on the infected population.

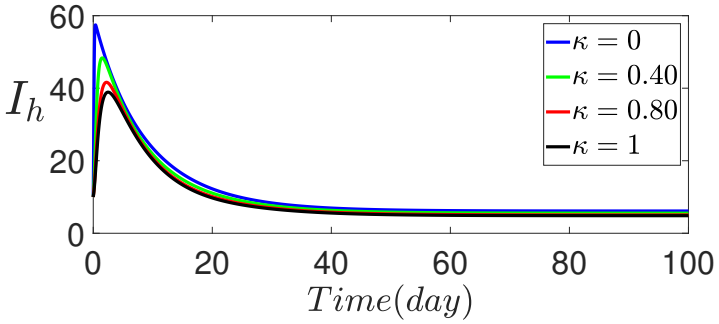


Figure 6: Influence of the media awareness parameter κ on the infected human population I_h.

Figure 6 illustrates the effect of media awareness on the infected human population I_h for different values of the media awareness effectiveness parameter κ. In the proposed model, media awareness acts through two mechanisms. First, the awareness level $M(t)$ promotes protective behavior by transferring susceptible individuals from S_h to the protected class P_h at the rate $\tau M S_h$. Second, it reduces the force of infection through the factor $\frac{1}{1+\kappa M}$. Therefore, increasing κ strengthens the effect of media awareness on reducing disease transmission. The numerical simulations consider $\kappa = 0$, 0.40, 0.80, and 1, while keeping the other parameters unchanged. As shown in Figure 6, increasing κ progressively reduces the number of infected individuals. When $\kappa = 0$, the infected population remains at a higher level, whereas larger values of κ lead to a

lower infection peak and a lower infection level over time. These results show numerically that increasing the effectiveness of media awareness can reduce the infection burden by promoting protective behavior and decreasing effective transmission. However, under the endemic conditions considered here, media awareness reduces the infection level but does not necessarily lead to complete disease elimination. Thus, media awareness represents an important behavioral control mechanism for reducing Hantavirus transmission.

F Sensitivity analysis of the basic reproduction number

This section is devoted to the sensitivity analysis of the model. The main objective is to determine the parameters that most strongly influence the dynamics of disease transmission and the basic reproduction number R_0. The following definition is required for the subsequent analysis.

Definition 1. [13] The normalized forward SI of the basic reproduction number R_0 with respect to a parameter θ measures the relative variation in R_0 resulting from a relative change in θ. It is defined by

$$S_\theta^{R_0} = \frac{\partial R_0}{\partial \theta}\frac{\theta}{R_0}.$$

Since $R_0 = \max\left\{R_0^h, R_0^r\right\}$, the sensitivity analysis is conducted separately for the two threshold quantities R_0^h and R_0^r. Specifically, we first calculate the normalized forward SIs of R_0^h with respect to the human-related parameters β_h, η, Λ_h, μ_h, and γ_h. We then determine the SIs of R_0^r with respect to the rodent-related parameters β_r, Λ_r, and μ_r.

The SIs of R_0^h are given by

$$\begin{cases} S_{\beta_h}^{R_0^h} = 1, \\ S_{\eta}^{R_0^h} = 1, \\ S_{\Lambda_h}^{R_0^h} = 1, \\ S_{\gamma_h}^{R_0^h} = -\dfrac{\gamma_h}{\gamma_h + \mu_h}, \\ S_{\mu_h}^{R_0^h} = -\dfrac{\gamma_h + 2\mu_h}{\gamma_h + \mu_h}. \end{cases} \tag{9}$$

The SIs of R_0^r are given by

$$\begin{cases} S_{\beta_r}^{R_0^r} = 1, \\ S_{\Lambda_r}^{R_0^r} = 1, \\ S_{\mu_r}^{R_0^r} = -2. \end{cases} \tag{10}$$

The computed SIs of the model parameters with respect to R_0^h and R_0^r are reported in Table 5 and illustrated in Figure 7.

Table 5: SIs of the model parameters associated with the basic reproduction numbers R_0^h and R_0^r.

Parameter	Value	SI for R_0^h	SI for R_0^r
Λ_h	1	1	–
β_h	0.25	1	–
η	0.80	1	–
γ_h	0.08	-0.4444	–
μ_h	0.10	-1.5556	–
Λ_r	1	–	1
β_r	0.20	–	1
μ_r	0.25	–	-2

Figure 7 and Table 5 show the normalized forward SIs of the model parameters with respect to the basic reproduction numbers R_0^h and R_0^r. In general, a positive SI indicates that an increase in the corresponding parameter leads to an increase in the basic reproduction number, whereas a negative SI indicates that increasing the parameter decreases the reproduction number. Furthermore, the magnitude of the SI reflects the relative importance of each parameter in influencing disease transmission. For the human basic reproduction number R_0^h, the parameters Λ_h, β_h, and η all have SIs equal to $+1$. Consequently, a 10% increase (or decrease) in any of these parameters results in a 10% increase (or decrease) in the value of R_0^h. This shows that the human recruitment rate, the human transmission rate, and the modification parameter have the largest positive impact on disease transmission within the human population. On the other hand, the SIs of the recovery rate γ_h and the natural death rate μ_h are both negative, equal to -0.4444 and -1.5556, respectively. Consequently, increasing either of these parameters decreases the reproduction number R_0^h, whereas decreasing them favors the spread of the disease. Similarly, for the rodent basic reproduction number R_0^r, the recruitment rate Λ_r and the transmission rate β_r have SIs of $+1$, indicating that a 10% increase in either parameter increases R_0^r by 10%. On the other hand, the natural death rate μ_r has an SI of -2, which means that a 10% increase in μ_r reduces R_0^r by approximately 20%. Among all rodent-related parameters,

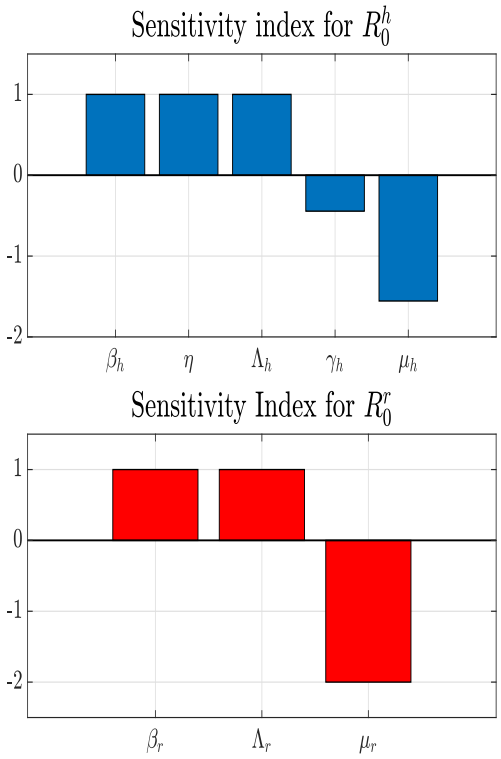


Figure 7: Normalized forward SIs of the parameters associated with R_0^h and R_0^r.

μ_r has the largest SI in absolute value and is therefore the most significant factor determining Hantavirus persistence in the rodent population. In summary, the sensitivity analysis shows that the transmission parameters β_h and β_r drive the spread of the disease, while the recovery rate γ_h and the natural death rates μ_h and μ_r help limit its spread. For this reason, reducing transmission through preventive actions, combined with measures aimed at speeding up recovery and shortening the infectious period of humans, is expected to be the most effective way to bring the basic reproduction numbers below one and thus limit the spread of Hantavirus among humans and rodents.

G Conclusion

This study proposed a novel fractional-order Hantavirus transmission model based on the Caputo fractional derivative that takes into account media awareness and the protective behavior of humans. By accounting for memory effects, the model better represents disease dynamics and

captures the complex interactions between humans and rodents. The mathematical analysis established the existence, uniqueness, positivity, and boundedness of the solutions, confirming that the model is mathematically well-posed and biologically meaningful. Moreover, the basic reproduction numbers were computed, and they determine the conditions for disease extinction or persistence.

The global stability analysis focused on two epidemiologically significant equilibria: The disease-free equilibrium E_0 and the endemic equilibrium E^*. While the human-only equilibrium E_h may exist when $R_0^h > 1$, it is of limited epidemiological significance, since Hantavirus is mainly sustained in infected rodent populations, the natural reservoir of the virus. Therefore, E_h does not properly describe the long-term dynamics of the coupled human-rodent transmission system. Similarly, the rodent-only equilibrium E_r, although biologically possible under certain circumstances, is not very informative about the public health impact of the disease because human infections are not involved. For this reason, a global stability analysis of E_h and E_r was not undertaken. Instead, only the disease-free and fully endemic equilibria were examined, since they determine the threshold behavior and the long-term dynamics of the complete transmission system. Specifically, it was shown that the disease-free equilibrium E_0 is globally asymptotically stable whenever $R_0 = \max\{R_0^h, R_0^r\} \leq 1$, whereas the endemic equilibrium E^* is globally asymptotically stable whenever $R_0^h > 1$ and $R_0^r > 1$. These global stability results were established using Lyapunov functions and LaSalle's invariance principle for fractional-order systems.

The simulations confirmed the theoretical results and illustrated the importance of the fractional-order parameter in the epidemic dynamics. In particular, reducing the fractional order accelerated convergence to the equilibrium without affecting its asymptotic stability. The simulations also showed that an increase in the transmission rates resulted in a higher number of infected individuals, whereas increasing the effectiveness of media awareness reduced the number of infections through changes in protective behavior. According to the sensitivity analysis, the transmission and recruitment parameters are the main contributors to the spread of the disease, while higher recovery and natural mortality rates are associated with a decrease in disease transmission. These findings highlight the importance of combining transmission-reduction measures, such as limiting contact between humans and infected rodents, with effective public awareness and appropriate protective behaviors as part of a comprehensive strategy for controlling Hantavirus transmission.

The proposed fractional-order framework provides a useful mathematical approach for investigating the combined effects of memory, media awareness, and behavioral responses on Hantavirus transmission dynamics. Future work could extend the present model to a multi-group setting to account for population heterogeneity and investigate optimal control strategies for reducing Hantavirus transmission [11].

Funding

This research received no external funding.

Conflicts of Interest

The author declares no conflicts of interest.

References

[1] Aamoum, I., Meskaf, A., Yaagoub, Z. and Farah, E.M., *Mathematical analysis of a fractional-order influenza model with resistant and non-resistant strains*, Iran. J. Numer. Anal. Optim. 16 (2026), 703–726.

[2] Abouchabane, M., Yaagoub, Z., Zamma, A. and Nissabouri, S., *A dynamic approach to transport logistics risks and resilience using a fractional SEIR model.* LogForum, 22 (3) (2026) 475–490,

[3] Alnafisah, Y. and El-Shahed, M., *Stochastic analysis of a Hantavirus infection model*, Mathematics **10** (2022), 3756.

[4] Anderson, R.M. and May, R.M., *Infectious diseases of humans: dynamics and control*, Oxford University Press, Oxford, 1991.

[5] Avsic-Zupanc, T., Saksida, A. and Korva, M., *Hantavirus infections*, Clin. Microbiol. Infect. 215 (2019), e6–e16.

[6] Bangia, A. and Bhardwaj, R., *Fractional ordered epidemiology model simulation for pathogenesis of the pandemic*, Franklin Open, 12 (2025), 100320.

[7] Batool, H., Khan, I., Li, W., Junaid, M., Zhang, J., Nawaz, A. and Tian, L., *Fractional modeling and numerical investigations of COVID-19 epidemic model with non-singular fractional derivatives: a case study*, Sci. Rep. 15 (2025), 13256.

[8] Bhatter, S., Kumawat, S., Purohit, S.D. and Suthar, D.L., *Mathematical modeling of tuberculosis using Caputo fractional derivative: a comparative analysis with real data*, Sci. Rep. 15 (2025), 12672.

[9] Brocato, R.L. and Hooper, J.W., *Progress on the prevention and treatment of hantavirus disease*, Viruses **11** (2019), 610.

[10] Davarian, N., Heydari, A., Mehne, S.H.H. and Heydari, A.A., *The investigation of stability and optimal control of the multigroup Covid-19 epidemic model in Iran*, J. Math. Ext. 19 (2025), 1–32.

[11] Davariyan, N., Heydari, A., Hashemi, M.S. and Heydari, A.A., *Presenting a multi-group model and optimal control of an infectious disease in Iran*, J. Oper. Research Appl. 22(1) (2025) 107–123 (in Persian).

[12] dos Santos, J.P.C., Monteiro, E. and Vieira, G.B., *Global stability of fractional SIR epidemic model*, Proc. Braz. Soc. Comput. Appl. Math. 5 (2017).

[13] El Bhih, A., Benfatah, Y., Hassouni, H., Balatif, O. and Rachik, M., *Mathematical modeling, sensitivity analysis, and optimal control of students awareness in mathematics education*, Partial Differ. Equ. Appl. Math. 11 (2024), 100795.

[14] Erturk, V.S., Momani, S. and Odibat, Z., *Application of generalized differential transform method to multi-order fractional differential equations*, Commun. Nonlinear Sci. Numer. Simul. 13 (2008), 1642–1654.

[15] Ghosh, B., *Fractional order modeling of ecological and epidemiological systems: ambiguities and challenges*, J. Anal. 33 (2025), 341–366.

[16] Gutiérrez-Jara, J.P., Muñoz-Quezada, M.T., Cordova-Lepe, F. and Silva-Guzmán, A., *Mathematical model of the spread of hantavirus infection*, Pathogens **12**(9) (2023), 1147.

[17] Jonsson, C.B., Figueiredo, L.T.M. and Vapalahti, O., *A global perspective on hantavirus ecology, epidemiology, and disease*, Clin. Microbiol. Rev. 23 (2010), 412–441.

[18] Keeling, M.J. and Rohani, P., *Modeling infectious diseases in humans and animals*, Princeton University Press, Princeton, 2008.

[19] Khumbudzo, M., Duah, E., Grobler, E. and Maluleke, K., *The role of mathematical modelling in predicting and controlling infectious disease outbreaks in underserved settings: a systematic review and meta-analysis*, Public Health Chall. 4 (2025), e70116.

[20] Kruger, D.H., Figueiredo, L.T.M., Song, J.W. and Klempa, B., *Hantaviruses: globally emerging pathogens*, J. Clin. Virol. 64 (2015), 128–136.

[21] Luis, A.D., Hayman, D.T., O'Shea, T.J., Cryan, P.M., Gilbert, A.T., Pulliam, J.R., Mills, J.N., Timonin, M.E., Willis, C.K., Cunningham, A.A. and Fooks, A.R., *A comparison of bats and rodents as reservoirs of zoonotic viruses: are bats special?*, Proc. R. Soc. B Biol. Sci. 280 (2013), 20122753.

[22] Moustafa, M., *Cost-effectiveness and optimal control of Hantavirus transmission in rodent populations*, Jambura J. Biomath. 7 (2026), 123–135.

[23] Moustafa, M., Hafiz Mohd, M., Izani Ismail, A. and Aini Abdullah, F., *Dynamical analysis of a fractional-order Hantavirus infection model*, Int. J. Nonlinear Sci. Numer. Simul. 21 (2020), 171–181.

[24] Prist, P.R., Uriarte, M., Tambosi, L.R., Prado, A., Pardini, R., D'Andrea, P.S. and Metzger, J.P., *Landscape, environmental and social predictors of Hantavirus risk in São Paulo, Brazil*, PLoS ONE **11** (2016), e0163459.

[25] Rida, S.Z., Abdel Rady, A.S., Arafa, A.A.M. and Khalil, M., *The effect of the environmental parameter on the Hantavirus infection through a fractional-order SI model*, Int. J. Basic Appl. Sci. 1(2), (2012) 88–99.

[26] Supriatna, A.K., Napitupulu, H., Ndii, M.Z., Ghosh, B. and Kon, R., *A mathematical model for transmission of hantavirus among rodents and its effect on the number of infected humans*, Comput. Math. Methods Med. 2023 (2023), 9578283.

[27] Tian, H. and Stenseth, N.C., *The ecological dynamics of hantavirus diseases: from environmental variability to disease prevention largely based on data from China*, PLoS Negl. Trop. Dis. 13 (2019), e0006901.

[28] Van den Driessche, P. and Watmough, J., *Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission*, Math. Biosci. 180 (2002), 29–48.

[29] Wang, Y.X., Voutilainen, L., Aminikhah, M., Helle, H., Huitu, O., Laakkonen, J. and Kallio, E.R., *The impact of wildlife and environmental factors on hantavirus infection in the host and its translation into human risk*, Proc. R. Soc. B Biol. Sci. 290 (2023), 20222470.

[30] Yaagoub, Z., *Fractional two-strain SVLIR epidemic model with vaccination and quarantine strategies*, Int. J. Dyn. Control **13** (2025), 55.

[31] Yaagoub, Z. and Ahmad, S., *Three-strain epidemic model for influenza virus involving fractional derivative and treatment*, J. Appl. Math. Comput. 71 (2025), 1247–1266.

[32] Yaagoub, Z., El Bhih, A. and Allali, K., *Global analysis of a fractional-order infection model for the propagation of computer viruses*, Model. Earth Syst. Environ. 11 (2025), 68.

[33] Yaagoub, Z., Sadki, M. and Allali, K., *A generalized fractional hepatitis B virus infection model with both cell-to-cell and virus-to-cell transmissions*, Nonlinear Dyn. 112 (2024), 16559–16585.

[34] Yakit Ongun, M. and Arslan, D., *Explicit and implicit schemes for fractional-order Hantavirus model*, Iran. J. Numer. Anal. Optim. 8 (2018), 75–93.

Iranian Journal of Numerical Analysis and Optimization
Vol. 17, No. 1, 2027, pp 603–628
https://doi.org/10.22067/ijnao.2026.96953.1800
https://ijnao.um.ac.ir/

CC BY NC

Research Article

An appropriate fractional Narayana polynomials neural network method for dental caries infection model

H. Hassani*, Z. Avazzadeh, A. Turan-Dincel and H. Farahmand

Abstract

Dental caries represents a major global health challenge, affecting individuals of all ages and from various socioeconomic backgrounds. Despite advancements in preventive measures and treatments, it remains the most prevalent chronic disease worldwide, severely impacting both oral and overall health. This condition arises from complex interactions between microorganisms, the host's biological makeup, diet, environmental factors, and behavioral patterns. These

*Corresponding author

Received 12 December 2025; revised 31 August 2026; accepted 2 September 2026

Hossein Hassani
Department of Mathematics, Anand International College of Engineering, Jaipur 303012, India. e-mail: hosseinhassani40@yahoo.com

Zakieh Avazzadeh
Stony Brook Institute at Anhui University, Anhui University, Hefei 230601, China. e-mail: z.avazzadeh@ahu.edu.cn

Arzu Turan-Dincel
Department of Mathematical Engineering, Yildiz Technical University, 34220, Esenler, Istanbul-Turkey. e-mail: artur@yildiz.edu.tr

Hanieh Farahmand
Department of Oral and Maxillofacial Medicine, School of Dentistry, Yasuj University of Medical Sciences, Yasuj, Iran. e-mail: haneyeh.farahmand@yums.ac.ir

How to cite this article
Hassani, H., Avazzadeh, Z., Turan-Dincel, A. and Farahmand, H., An appropriate fractional Narayana polynomials neural network method for dental caries infection model. *Iran. J. Numer. Anal. Optim.*, 2027; 17(1): 603-628. https://doi.org/10.22067/ijnao.2026.96953.1800

factors collectively contribute to the gradual demineralization and breakdown of tooth structures, leading to cavities and tooth decay. To address the mathematical model for simulating the spread of dental caries, a more accurate and effective neural network-based optimization (NNO) technique is suggested. In the present contribution, we introduce new fractional Narayana polynomials (FNPs). The NNO method consists of three main layers, which include an input layer, a hidden layer, and an output layer. The time variable (in months) is employed as an input parameter to the NNO method. Within the structure of the neural network, FNPs are employed as activation functions in the hidden layer, whereas the $arcsinh(t)$ function is utilized for the output layer. The mathematical model for dental caries is reduced to the problem of solving a system of algebraic equations through the use of FNPs, NNO and the Lagrange multiplier method. The convergence analysis, existence and uniqueness of solutions are discussed. The validity of the proposed NNO method is confirmed through relevant simulations. With its high efficiency and accuracy, the proposed NNO method proves to be an effective tool for analyzing the mathematical model for dental caries, showing significant potential for applications in two-dimensional differential equations and optimal control problems.

AMS subject classifications (2020): 41A10; 65Z05; 92C42.

Keywords: Fractional Narayana polynomials neural network; Dental caries infection model; Optimization algorithm; Lagrange multiplier method.

A Introduction

The development of dental caries is driven by an imbalance between the demineralization and remineralization processes in the enamel. Acid-producing bacteria within plaque biofilm metabolize fermentable carbohydrates such as glucose, sucrose, and fructose into acids that erode the enamel and dentin, initiating caries formation [10]. As a result, dental caries is considered a microbial dietary disease, requiring both a cariogenic biofilm and frequent exposure to fermentable sugars [18, 43]. This continuous interaction between diet and microorganisms plays a crucial role in the progression of the disease [24]. Host susceptibility significantly influences the onset and progression of dental caries. Genetic factors affect immune responses, enamel structure, and saliva production, all of which impact an individual's resistance to cariogenic bacteria [35]. Salivary properties including pH, buffering capacity, and antimicrobial components help neutralize acids and regulate bacterial growth [25]. Additionally, fluoride exposure strengthens enamel and inhibits bacterial metabolism, making it a key factor in caries prevention. However, disparities in access to fluoridated water and dental products contribute to variations in caries prevalence across different populations [37]. Beyond biological and dietary influences, environmental and

behavioral factors also increase the risk of dental caries. Socioeconomic status affects access to dental care, oral health education, and fluoride-based products [41]. Poor oral hygiene promotes bacterial biofilm accumulation, leading to prolonged acid exposure and accelerated enamel demineralization, thereby increasing the risk of caries [36]. Furthermore, frequent consumption of fermentable carbohydrates creates an acidic oral environment, promoting both enamel breakdown and bacterial biofilm formation [27]. Several key risk factors contribute to dental caries: a history of caries, plaque accumulation, inadequate fluoride exposure, diet, reduced salivary flow, and poor overall health status [12]. Untreated dental caries in permanent teeth remains a widespread health issue, emphasizing the urgent need for improved prevention strategies and early interventions [17].

In recent years, mathematical models have emerged as a powerful tool for representing complex biological systems. A large number of numerical and neural network (NN) methods have been carried out in practical fields such as mathematical models of diverse diseases. Li and Yang [22] applied a numerical method based on the linearly implicit Euler method for an HIV model with latently infected T cells. Wang et al. [42] discussed the numerical solution of an HIV model by employing the linearly implicit Euler method with a general transmission dynamic. He [14] proposed a machine learning technique to obtain the solution of the HIV model of CD4(+). Sabir, Ben Said, and Al-Mdallal [33] investigated the numerical performances of the fractional-order mathematical influenza disease model by designing the computational framework based on the stochastic Levenberg–Marquardt backpropagation NNs. Moreover, Sabir et al. [32] presented the numerical solutions of the chickenpox disease model by applying a proficient optimization framework based on the artificial NN. Anusha Bamini et al. [3] used the convolutional NN in deep learning technique for premature detection of Alzheimer's disease. Shafik et al. [34] presented a hybrid inception-xception using a convolution NN for efficient plant disease classification and detection. Chen, Yang, and Sheng, [9] studied the numerical analysis of linearly implicit Euler method for age-structured SIS model. Rafique et al. [29] formulated the hybridized finite difference numerical method for obtaining the numerical solutions of the delayed fractional diabetes mellitus model. Avazzadeh et al. [4] presented an optimization method using generalized Laguerre polynomials and Lagrange multipliers to study a fractional tuberculosis model. Priyanga, Pattankar, and Sridevi [28] proposed a hybrid model using a recurrent NN and logistic chaos-based whale optimization for heart disease prediction from electronic health records. Iqbal et al. [16] presented the reciprocal domain adaptation network to the various issues of chronic kidney disease diagnosis. Amit Prakash [2] used the fractional Adams-Bashforth approach to examine the childhood disease. Dhar and Gupta [11] proposed a multi-step differential transformation method for solving tumor-immune model with B cells and monoclonal antibody drug. Avazzadeh et al. [5] presented an optimized method based on generalized Lerch polynomials to solve the

fractional-order model of CAR T-cell therapy for T-cell leukemia. Az-Zobi et al. [6] used the modified physics informed NNs with impacts of fractional derivative to obtain numerical solutions of modified Atangana Baleanu Caputo fractional advection dispersion models. Az-Zobi et al. [7] applied a feed-forward NN with five fully connected layers trained on data samples generated from the exact solutions of specific ordinary differential equations. Ruttanaprommarin et al. [31] applied a stochastic framework based on Levenberg–Marquardt backpropagation NNs for the Holling third kind of functional response delay differential model.

The current work introduces a new methodology to approximate the optimal solutions of susceptible human population $S(t)$ due to dental caries, the exposed bot not yet infected human population, $E(t)$, infected human population $I(t)$, and the recovered human population through the treatment $R(t)$ functions at time t, $t \in [0,T]$ using a neural network-based optimization (NNO) method and new fractional Narayana polynomials (FNPs). The mathematical model for simulating the spread of dental caries having the ordinary differential equation is given by [19]

$$\begin{cases} \frac{dS(t)}{dt} = \Lambda + \gamma R(t) - \left(\mu + \frac{\beta I(t)}{1+aI(t)}\right) S(t), \\ \frac{dE(t)}{dt} = \frac{\beta S(t) I(t)}{1+aI(t)} - (m+\rho)\, E(t), \\ \frac{dI(t)}{dt} = \rho E(t) - \left(\mu + \xi + \omega + \frac{\nu}{1+\eta I(t)}\right) I(t), \\ \frac{dR(t)}{dt} = \left(\omega + \frac{\nu}{1+\eta I(t)}\right) I(t) - (\mu+\gamma)\, R(t), \\ S(0) = S_0 \geq 0, \quad E(0) = E_0 \geq 0, \quad I(0) = I_0 \geq 0, \quad R(0) = R_0 \geq 0. \end{cases} \tag{1}$$

A description of the parameters in model (1) is provided in Table 1. The primary objectives of this research are as follows:

(i) This study is the first to present the concept of FNPs, a generalization of Narayana polynomials (NPs).

(ii) The convergence analysis, existence and uniqueness of solutions are provided for the method.

(iii) An optimization algorithm is derived using the Lagrange multiplier method.

(iv) A few terms of the FNPs are utilized to attain highly accurate results.

(v) The considered problem is transformed into a system of algebraic equations, which can be solved using an appropriate numerical method.

(vi) The proposed approach is computer-based and easy to apply in any language that supports programming.

(vii) The accuracy of the method is shown by solving numerical examples.

The paper is organized as follows: We present a description of NPs in Section B. Section C presents a brief overview of the FNPs. Section D discusses the application of the suggested

Table 1: Definition of parameters in model (1) [19].

Parameter	Description
Λ	rate of recruitment of the population
γ	rate at which for recovered individuals move back to susceptible class
μ	natural death rate
β	the coefficient of dental infection transmission
a	the parameter that measures the inhibitory effects
m	mortality rate of the population and cut across all the compartments
ρ	rate at which exposed individuals became infected
ξ	death rate due to the infection
ω	recovery rate of infected individuals due to the response
ν	maximum medical resources supplied per unit time
η	A saturated factor that tracks the impact of delayed therapy on infected humans

NNO method to the given model. Sections E and F provide a convergence analysis of the method and existence and uniqueness of solutions for the system (1), respectively. To verify the theoretical results, Section G presents simulation results and applied examples. These simulations demonstrate the influence of various parameters on the model. Section H outlines the main conclusions of the paper and discusses their implications for future work.

B NPs and function approximation

In this part of the paper, we discuss the definitions of NPs and the concept of function approximation using NPs.

Definition 1. Narayana numbers $\mathcal{N}_{i,j}$ are defined by the relations [26, 1]:

$$\mathcal{N}_{0,j} = \mathcal{N}_{j,0} = \delta_{j0} \quad (j \in \mathbb{N}_0), \qquad \mathcal{N}_{i,j} = \frac{1}{i}\binom{i}{j}\binom{i}{j-1} \quad (i, j \in \mathbb{N}),$$

where δ_{ij} denotes the Kronecker delta function.

Definition 2. The analytical form of the NPs, $\mathbb{A}_n(t)$, is obtained by [26, 1]

$$\mathbb{A}_n(t) = \sum_{k=0}^{n} \mathcal{N}_{n+1,k+1} t^k = \sum_{k=0}^{n} \frac{1}{n+1}\binom{n+1}{k+1}\binom{n+1}{k} t^k, \qquad n \in \mathbb{N}_0.$$

The first NPs are given by $\mathbb{A}_0(t) = 1$, $\mathbb{A}_1(t) = 1+t$, $\mathbb{A}_2(t) = 1+3t+t^2$, $\mathbb{A}_3(t) = 1+6t+6t^2+t^3$, and $\mathbb{A}_4(t) = 1+10t+20t^2+10t^3+t^4$.
A function $f(t)$ can be approximated using the first $(n+1)$ terms of the NPs as follows:

$$f(t) \simeq Q_n^T R_n S_n(t),$$

where $Q_n^T = [q_0 \;\; q_1 \;\; \dots \;\; q_n]$, and

$$R_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ r_{1,0} & r_{1,1} & \cdots & r_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n,0} & r_{n,1} & \cdots & r_{n,n} \end{pmatrix}, \quad S_n(t) = [1 \;\; t \;\; t^2 \;\; \dots \;\; t^n]^T,$$

and

$$r_{ij} = \begin{cases} \dfrac{1}{i+1}\binom{i+1}{j+1}\binom{i+1}{j}, & i \geq j, \\ 0, & i < j. \end{cases}$$

C FNPs and NN method

The definition of FNPs, their use in function approximation, and the Fractional Narayana polynomials neural network (FNPNN) model are introduced in this section.

Definition 3. The FNPs, $\mathbf{A}_m(t)$, are defined by the change of variable t^k into $t^{k+\alpha_k}$, $(k+\alpha_k > 0)$ on NPs as follows:

$$\mathbf{A}_m(t) = \sum_{k=0}^{m} \frac{1}{m+1}\binom{m+1}{k+1}\binom{m+1}{k} t^{k+\alpha_k}, \qquad m \in \mathbb{N}_0,$$

where α_k denotes a control parameter. When $\alpha_k = 0$, FNPs and classical NPs coincide.

C.1 Function approximations

A matrix representation of the functions $S(t)$, $E(t)$, $I(t)$ and $R(t)$ in (1) is provided below:

$$S(t) \simeq \mathbf{B}^T \mathbf{U} \mathbf{W}_{m_1}(t), \qquad E(t) \simeq \mathbf{C}^T \mathbf{V} \mathbf{W}_{m_2}(t),$$
$$I(t) \simeq \mathbf{D}^T \mathbf{P} \mathbf{W}_{m_3}(t), \qquad R(t) \simeq \mathbf{L}^T \mathbf{Q} \mathbf{W}_{m_4}(t),$$

where

$$\mathbf{B}^T = [b_0 \;\; b_1 \dots b_{m_1}], \quad \mathbf{C}^T = [c_0 \;\; c_1 \dots c_{m_2}],$$
$$\mathbf{D}^T = [d_0 \;\; d_1 \dots d_{m_3}], \quad \mathbf{L}^T = [l_0 \;\; l_1 \dots l_{m_4}], \tag{2}$$

$$\mathbf{U} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ u_{1,0} & u_{1,1} & u_{1,2} & \cdots & u_{1,m_1} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ u_{m_1,0} & u_{m_1,1} & u_{m_1,2} & \cdots & u_{m_1,m_1} \end{pmatrix},$$
$$\mathbf{V} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ v_{1,0} & v_{1,1} & v_{1,2} & \cdots & v_{1,m_2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ v_{m_2,0} & v_{m_2,1} & v_{m_2,2} & \cdots & v_{m_2,m_2} \end{pmatrix}, \tag{3}$$

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ p_{1,0} & p_{1,1} & p_{1,2} & \cdots & p_{1,m_3} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ p_{m_3,0} & p_{m_3,1} & p_{m_3,2} & \cdots & p_{m_3,m_3} \end{pmatrix},$$
$$\mathbf{Q} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ q_{1,0} & q_{1,1} & q_{1,2} & \cdots & q_{1,m_4} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ q_{m_4,0} & q_{m_4,1} & q_{m_4,2} & \cdots & q_{m_4,m_4} \end{pmatrix}, \tag{4}$$

$$u_{ij} = \begin{cases} \dfrac{1}{i+1}\binom{i+1}{j+1}\binom{i+1}{j}, & i \geq j, \\ 0, & i < j. \end{cases} \qquad i = 1, 2, \dots, m_1, \;\; j = 0, 1, \dots, m_1,$$

$$v_{ij} = \begin{cases} \dfrac{1}{i+1}\binom{i+1}{j+1}\binom{i+1}{j}, & i \geq j, \\ 0, & i < j. \end{cases} \qquad i = 1, 2, \ldots, m_2, \quad j = 0, 1, \ldots, m_2,$$

$$p_{ij} = \begin{cases} \dfrac{1}{i+1}\binom{i+1}{j+1}\binom{i+1}{j}, & i \geq j, \\ 0, & i < j. \end{cases} \qquad i = 1, 2, \ldots, m_3, \quad j = 0, 1, \ldots, m_3,$$

$$q_{ij} = \begin{cases} \dfrac{1}{i+1}\binom{i+1}{j+1}\binom{i+1}{j}, & i \geq j, \\ 0, & i < j. \end{cases} \qquad i = 1, 2, \ldots, m_4, \quad j = 0, 1, \ldots, m_4,$$

and

$$\mathbf{W}_{m_i}(t) \triangleq [w_0^i(t) \;\; w_1^i(t) \ldots w_{m_i}^i(t)]^T, \qquad i = 1, 2, 3, 4,$$

$$w_j^i(t) = \begin{cases} 1, & j = 0, \\ t^{j+\zeta_j^i}, & j = 1, 2, \ldots, m_i, \end{cases} \qquad i = 1, 2, 3, 4, \tag{5}$$

with ζ_j^i representing the control parameters as follows:

$$\mathfrak{Q}^1 = [\zeta_1^1 \;\; \zeta_2^1 \ldots \zeta_{m_1}^1], \quad \mathfrak{Q}^2 = [\zeta_1^{21} \;\; \zeta_2^2 \ldots \zeta_{m_2}^2],$$
$$\mathfrak{Q}^3 = [\zeta_1^3 \;\; \zeta_2^3 \ldots \zeta_{m_3}^3], \quad \mathfrak{Q}^4 = [\zeta_1^4 \;\; \zeta_2^4 \ldots \zeta_{m_4}^4].$$

A description of parameters in this section is provided in Table 2.

C.2 FNPNN

NNs are known as effective machine learning algorithms that are ideal for solving complicated problems [39, 38, 40]. In order to efficiently and accurately solve a mathematical model of dental caries infection, a novel FNPNN method is proposed in this paper. The FNPNN method is structured with three layers that function as described below:

(i) The input t is provided by the network's input layer.

Table 2: Definition of parameters in Section 3.

Parameter	Description
t	Time
$\mathbb{A}_n(t)$	NPs
$\mathbf{A}_m(t)$	FNPs
$\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$, $\mathbf{L}$	Vectors of free coefficients
ζ_j^i	Control parameters
$\mathbf{W}_{m_1}$, $\mathbf{W}_{m_2}$, $\mathbf{W}_{m_3}$, $\mathbf{W}_{m_4}$	Vectors of control parameters
$\mathbf{U}$, $\mathbf{V}$, $\mathbf{P}$, $\mathbf{Q}$	Matrices of NP coefficients
$\mathfrak{Q}^i(t)$	Vectors of FNPs
$\tilde{\mathbf{S}}(t)$, $\tilde{\mathbf{E}}(t)$, $\tilde{\mathbf{I}}(t)$, $\tilde{\mathbf{R}}(t)$	Approximations of $\mathbf{S}(t)$, $\mathbf{E}(t)$, $\mathbf{I}(t)$ and $\mathbf{R}(t)$

(ii) In the network's hidden layer, we select FNPs of varying degrees as activation functions.
(iii) The network's output layer is constructed as a weighted linear combination of FNPs. Figure 1 illustrates the NN structure utilized in this study. Note that the activation function is the function that acts upon the input to get the output of the network. There are many types of activation functions such as the sigmoid function, the hyperbolic tangent function, the ReLu function and others [8]. In this context, the inverse hyperbolic sine (arcsine) functions are used as activation functions. Consequently, the following are the results of the FNPNN, using input data t and parameters $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$ and $\mathbf{L}$:

$$\begin{aligned} \mathbf{S}(t,\mathbf{B}) &= arc\sinh(S(t)), \qquad \mathbf{E}(t,\mathbf{C}) = arc\sinh(E(t)),\\ \mathbf{I}(t,\mathbf{D}) &= arc\sinh(I(t)), \qquad \mathbf{R}(t,\mathbf{L}) = arc\sinh(R(t)), \end{aligned} \tag{6}$$

where $S(t)$, $E(t)$, $I(t)$ and $R(t)$ are linear combinations expressed as follows:

$$\begin{aligned} S(t) &= \sum_{i=0}^{m_1} b_i \mathbf{A}_i(t), \qquad E(t) = \sum_{j=0}^{m_2} c_j \mathbf{A}_j(t),\\ I(t) &= \sum_{k=0}^{m_3} d_k \mathbf{A}_k(t), \qquad R(t) = \sum_{p=0}^{m_4} l_k \mathbf{A}_p(t), \end{aligned}$$

and $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$ and $\mathbf{L}$ vectors are given by

$$\mathbf{B}^T = [b_0 \;\; b_1 \ldots b_{m_1}], \qquad \mathbf{C}^T = [c_0 \;\; c_1 \ldots c_{m_2}],$$

$$\mathbf{D}^T = [d_0 \;\; d_1 \dots d_{m_3}], \qquad \mathbf{L}^T = [l_0 \;\; l_1 \dots l_{m_4}].$$

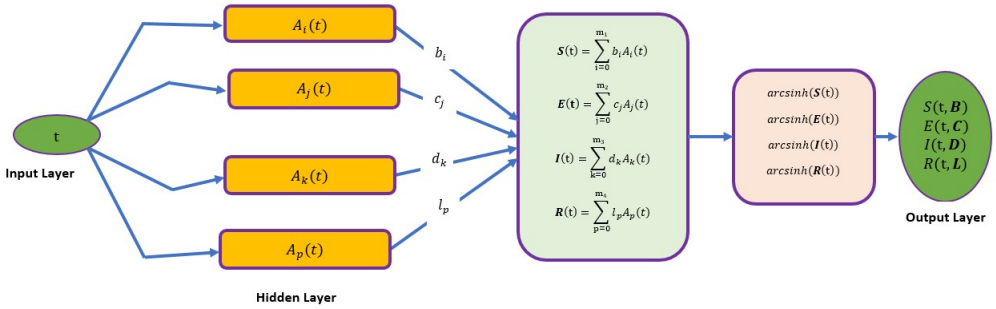


Figure 1: FNPNN architecture.

D Description of the approach

The focus of this section is to introduce the FNPNN approach for solving the dental caries infection model in (1). We approximate $S(t)$, $E(t)$, $I(t)$ and $R(t)$ by using the activation functions as

$$\begin{aligned} S(t) &\simeq \tilde{S}(t) = S_0 + t\mathbf{S}(t, \mathbf{B}), & E(t) &\simeq \tilde{E}(t) = E_0 + t\mathbf{E}(t, \mathbf{C}), \\ I(t) &\simeq \tilde{I}(t) = I_0 + t\mathbf{I}(t, \mathbf{D}), & R(t) &\simeq \tilde{R}(t) = R_0 + t\mathbf{R}(t, \mathbf{L}). \end{aligned} \tag{7}$$

We know that $S(t)$, $E(t)$, $I(t)$ and $R(t)$ satisfy the initial conditions. By applying (7) to the initial conditions of problem (1), we derive the following properties:

$$\begin{aligned} \Delta_1 &\triangleq \tilde{S}(0) - S_0 \simeq 0, & \Delta_2 &\triangleq \tilde{E}(0) - E_0 \simeq 0, \\ \Delta_3 &\triangleq \tilde{I}(0) - I_0 \simeq 0, & \Delta_4 &\triangleq \tilde{R}(0) - R_0 \simeq 0. \end{aligned} \tag{8}$$

Using the approximations of (7) in (1), we have

$$\begin{cases} \frac{d\tilde{S}(t)}{dt} - \Lambda - \gamma\tilde{R}(t) + \left(\mu + \frac{\beta\tilde{I}(t)}{1+a\tilde{I}(t)}\right)\tilde{S}(t) \triangleq \mathscr{R}_1(t, \mathbf{B}, \mathbf{W}^1), \\ \frac{d\tilde{E}(t)}{dt} - \frac{\beta\tilde{S}(t)\tilde{I}(t)}{1+a\tilde{I}(t)} + (m+\rho)\,\tilde{E}(t) \triangleq \mathscr{R}_2(t, \mathbf{C}, \mathbf{W}^2), \\ \frac{d\tilde{I}(t)}{dt} - \rho\tilde{E}(t) + \left(\mu + \xi + \omega + \frac{\nu}{1+\eta\tilde{I}(t)}\right)\tilde{I}(t) \triangleq \mathscr{R}_3(t, \mathbf{D}, \mathbf{W}^3), \\ \frac{d\tilde{R}(t)}{dt} - \left(\omega + \frac{\nu}{1+\eta\tilde{I}(t)}\right)\tilde{I}(t) + (\mu+\gamma)\,\tilde{R}(t) \triangleq \mathscr{R}_4(t, \mathbf{L}, \mathbf{W}^4). \end{cases} \tag{9}$$

Here, $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$ and $\mathbf{L}$ denote unknown free coefficients. Additionally, $\mathbf{W}^i$, $(i = 1, 2, 3, 4)$, are unknown control parameters, which are defined as follows:

$$\mathbf{W}^i = \left[\zeta_1^i \;\; \zeta_2^i \;\; \ldots \;\; \zeta_{m_i}^i\right], \qquad i = 1, 2, 3, 4.$$

The 2-norm of the residual functions is given by

$$\begin{aligned} &\mathfrak{R}\left(\mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{L}, \mathbf{W}^1, \mathbf{W}^2, \mathbf{W}^3, \mathbf{W}^4\right) \\ &= \int_0^T \left(\mathscr{R}_1^2(t, \mathbf{B}, \mathbf{W}^1) + \mathscr{R}_2^2(t, \mathbf{C}, \mathbf{W}^2) + \mathscr{R}_3^2(t, \mathbf{D}, \mathbf{W}^3) + \mathscr{R}_4^2(t, \mathbf{L}, \mathbf{W}^4)\right) dt. \end{aligned} \tag{10}$$

It is required to evaluate the unknown vectors $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$, $\mathbf{L}$, $\mathbf{W}^1$, $\mathbf{W}^2$, $\mathbf{W}^3$, and $\mathbf{W}^4$ by employing the optimization problem as follows:

$$\min\, \mathfrak{R}\left(\mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{L}, \mathbf{W}^1, \mathbf{W}^2, \mathbf{W}^3, \mathbf{W}^4\right), \tag{11}$$

under the conditions in (8) where $\mathfrak{R}$ is the objective function. To address this optimization problem, we assume the following condition:

$$\begin{aligned} &\mathcal{J}^*[\mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{L}, \mathbf{W}^1, \mathbf{W}^2, \mathbf{W}^3, \mathbf{W}^4, \theta] \\ &= \mathfrak{R}\left(\mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{L}, \mathbf{W}^1, \mathbf{W}^2, \mathbf{W}^3, \mathbf{W}^4\right) + \theta\,\Delta. \end{aligned} \tag{12}$$

Note that θ is the Lagrange multipliers vector. By utilizing the Lagrange multiplier technique, we can derive the necessary and sufficient conditions for optimality, which are as follows:

$$\begin{cases} \dfrac{\partial\mathcal{J}^*}{\partial\theta} = 0, & \dfrac{\partial\mathcal{J}^*}{\partial\mathbf{B}} = 0, & \dfrac{\partial\mathcal{J}^*}{\partial\mathbf{C}} = 0, \\ \dfrac{\partial\mathcal{J}^*}{\partial\mathbf{D}} = 0, & \dfrac{\partial\mathcal{J}^*}{\partial\mathbf{L}} = 0, & \dfrac{\partial\mathcal{J}^*}{\partial\mathbf{W}^i} = 0, \qquad i = 1, 2, 3, 4. \end{cases} \tag{13}$$

By solving that system for the unknown vectors $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$, $\mathbf{L}$, $\mathbf{W}^1$, $\mathbf{W}^2$, $\mathbf{W}^3$, and $\mathbf{W}^4$, we attain the approximate optimal solution for (1). In this study, the nonlinear algebraic system

given in (13) is solved using the 'fsolve' tool in Maple 2024. The procedure above outlines the application of the FNPNN method to solve the dental caries infection model presented in (1). The step-by-step algorithm and flowchart of the explained approach are concisely stated below:

FNPNN Algorithm for solving a mathematical model of the dental caries infection in (1)

Input: Initial conditions, Number of basis functions (m_i, $i = 1, 2, 3, 4$).

* Determine the unknown vectors $\mathbf{B}$, $\mathbf{C}$, $\mathbf{D}$, $\mathbf{L}$ and matrices of NPs coefficients $\mathbf{U}$, $\mathbf{V}$, $\mathbf{P}$ and $\mathbf{Q}$ based on (2)–(4).
* Determine the basis functions $w_j^i(t)$ by (5).
* Determine the activation functions $\mathbf{S}(t, \mathbf{B})$, $\mathbf{E}(t, \mathbf{C})$, $\mathbf{I}(t, \mathbf{D})$ and $\mathbf{R}(t, \mathbf{L})$ using (6).
* Calculate:
  . The initial conditions Δ_i for $i = 1, 2, 3, 4$ using (8).
  . The residual functions $\mathscr{R}_i$ for $i = 1, 2, 3, 4$ using (9).
  . The 2-norm of the residual functions $\mathfrak{R}\left(\mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{L}, \mathbf{W}^1, \mathbf{W}^2, \mathbf{W}^3, \mathbf{W}^4\right)$ using (10).
* Minimize the objective function $\mathfrak{R}\left(\mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{L}, \mathbf{W}^1, \mathbf{W}^2, \mathbf{W}^3, \mathbf{W}^4\right)$ under the constraints defined in (11).
* Apply the Lagrange multiplier technique $\mathcal{J}^*[\mathbf{B}, \mathbf{C}, \mathbf{D}, \mathbf{L}, \mathbf{W}^1, \mathbf{W}^2, \mathbf{W}^3, \mathbf{W}^4, \theta]$ using (12).
* Solve the nonlinear algebraic system using (13).

Output: The approximate solutions are $S(t) \simeq \tilde{S}(t) = S_0 + t\mathbf{S}(t, \mathbf{B})$, $E(t) \simeq \tilde{E}(t) = E_0 + t\mathbf{E}(t, \mathbf{C})$, $I(t) \simeq \tilde{I}(t) = I_0 + t\mathbf{I}(t, \mathbf{D})$, $R(t) \simeq \tilde{R}(t) = R_0 + t\mathbf{R}(t, \mathbf{L})$.

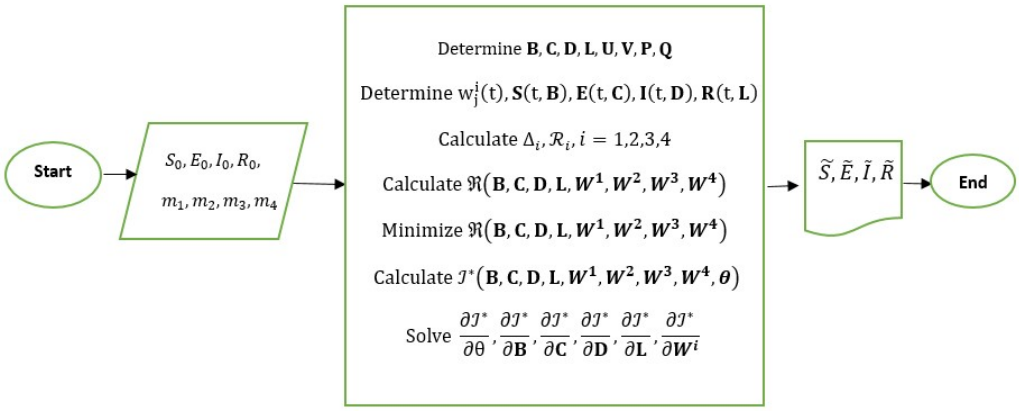


Figure 2: The flowchart of the explained approach.

E Convergence analysis

In this section, we provide theorems that guarantee the convergence of the proposed method.

Theorem 1. Consider (e_n) is a sequence in X such that $Span\{\mathrm{e}_n | n \in \mathbb{N}\}$ is dense in X; $(X, \|\cdot\|)$ is a normed linear space; and $x_0 \in X$. Let (x_n) be the best approximation to x_0 in the $Span\{\mathrm{e}_1, \ldots, \mathrm{e}_n\}$; then (x_n) converges to x_0.

Proof. Since the closure of $Span\{e_n | n \in \mathbb{N}\}$ is equal to X, there exists a sequence (y_k) in $Span\{e_n | n \in \mathbb{N}\}$ with $y_k \to x_0$. We suppose that $y_k \in Span\{e_1, \ldots, e_{n_k}\}$, in which (n_k) is a monotonically increasing sequence in $\mathbb{N}$, without losing generality. Therefore, we can write

$$\|x_{n_k} - x_0\| \leq \|y_k - x_0\| \to 0, \quad (\text{as } k \to \infty). \tag{14}$$

It is also evident that the sequence $(\|x_n - x_0\|)_{n\in\mathbb{N}}$, consisting of non-negative real numbers is non-increasing. By applying (14), this sequence contains a convergent subsequence. As a result, (x_n) converges to x_0. □

Theorem 2. Let $g : [0, 1] \to \mathbb{R}$ be a continuous function. Then, for every $t \in (0, 1)$ and $\epsilon > 0$, there exists an FNPs, $\mathbf{A}_m(t)$, such that

$$|g(t) - \mathbf{A}_m(t)| < \epsilon.$$

Proof. Let $\epsilon > 0$ be arbitrarily chosen. In view of the Weierstrass theorem [20], there exists a polynomial $P_m(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_m t^m$, $t \in (0, 1)$, such that

$$\sup_{t\in(0,1)} |g(t) - P_m(t)| < \frac{\epsilon}{2}.$$

We construct a FNPs, $\mathbf{A}_m(t)$, as follows:

Let $t \in (0, 1)$. We consider a sequence $\{\gamma_{k,n}\}_{n\in\mathbb{N}}$ of real numbers such that $\rho_{k,n} \to \frac{-\ln\left(\binom{m}{k}(1-t)^{m-k}\right)}{\ln(t)}$ as $n \to \infty$, for all $k = 0, 1, 2, 3, \ldots, m$, where $\binom{m}{k} = \frac{m!}{k!(m-k)!}$. This ensures that $t^{\rho_{k,n}} \to \frac{1}{c_{mk}}$ as $n \to \infty$, for all $k = 0, 1, 2, 3, \ldots, m$, where

$$c_{mk} = \frac{-\ln\left(\binom{m}{k}(1-t)^{m-k}\right)}{\ln(t)}.$$

Hence, for every $\epsilon > 0$, there exist $N_0, N_1, \ldots, N_m$ in $\mathbb{N}$ such that

$$|1 - c_{mk} t^{\rho_{k,n}}| < \frac{\epsilon}{2\max\{|a_0|, |a_1|, \ldots, |a_m|\}(m+1)}, \quad \forall\ k = 0, 1, 2, 3, \ldots, m.$$

Setting $N = \max\{N_0, N_1, \ldots, N_m\}$ and

$$\mathbf{A}_m(t) = a_0 c_{m0} t^{\rho_{0,N}} + a_1 c_{m1} t^{1+\rho_{1,N}} + a_2 c_{m2} t^{2+\rho_{2,N}} + \cdots + a_m c_{mm} t^{m+\rho_{m,N}},$$

we conclude that

$$|P_m(t) - \mathbf{A}_m(t)| < \frac{\epsilon}{2}.$$

This implies that

$$|g(t) - \mathbf{A}_m(t)| \leq |g(t) - P_m(t)| + |P_m(t) - \mathbf{A}_m(t)| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

□

Remark 1. For each $T > 0$, we consider $X = L^2[0, T]$, equipped with $\|\cdot\|_2$, and $\mathrm{e}_n := \mathbf{A}_n$, the FNPs. Then, by Theorem 1, it follows that for every $x_0 \in L^2[0, T]$ the sequence (x_n) of best approximation of x_0 in $Span\{\mathbf{A}_1, \ldots, \mathbf{A}_n\}$, converges (in $\|\cdot\|_2$) to x_0.

Remark 2. Let $\tilde{S}(t)$, $\tilde{E}(t)$, $\tilde{I}(t)$ and $\tilde{R}(t)$ be the FNPNN approximate solutions of $S(t)$, $E(t)$, $I(t)$ and $R(t)$, respectively. We refer the readers to [30] for the error bound of approximations.

F Existence and uniqueness of solutions for the system (1)

Lemma 1. [21, 23] Let $\Omega \subseteq \mathbb{R}^n$, and consider the differential system

$$\dot{x}(t) = f(t, x), \quad (t > 0),$$

with initial condition x_0, where $f : [0, \infty) \times \Omega \to \mathbb{R}^n$ satisfies a local Lipschitz condition with respect to x on $[0, \infty) \times \Omega$. Then this equation admits a unique solution $x(t)$.

Theorem 3. For $0 < t < \infty$, the system of differential equations (1) has a unique solution $\big(S(t), E(t), I(t), R(t)\big)$.

Proof. Let $F : [0, \infty) \times \mathbb{R}^4 \to \mathbb{R}^4$ be defined for each $t \in [0, \infty)$, $a \neq 0$, $1 + aI(t) > 0$, $1 + \eta I(t) > 0$ and $X = (S, E, I, R) \in \mathbb{R}^4$ by $F(t, X) = F(t, S, E, I, R) = \big(f_1(X), f_2(X), f_3(X), f_4(X)\big)$, where

$$\begin{aligned}
f_1(X) &= \Lambda + \gamma R(t) - \left(\mu + \tfrac{\beta I(t)}{1+aI(t)}\right) S(t), \\
f_2(X) &= \tfrac{\beta S(t) I(t)}{1+aI(t)} - (m + \rho)\, E(t), \\
f_3(X) &= \rho E(t) - \left(\mu + \xi + \omega + \tfrac{\nu}{1+\eta I(t)}\right) I(t), \\
f_4(X) &= \left(\omega + \tfrac{\nu}{1+\eta I(t)}\right) I(t) - (\mu + \gamma)\, R(t).
\end{aligned} \tag{15}$$

In order to prove the uniqueness of the solution for system (1), it suffices, according to Lemma 1, to demonstrate that F is locally Lipschitz in the region $[0,T]\times\mathcal{B}$, where M and T are both arbitrary positive real numbers and $\mathcal{B}:=\{X\in\mathbb{R}^4:\ \|X\|_\infty\leq M\}$. Now, let $X=(S,E,I,R),\tilde{X}=(\tilde{S},\tilde{E},\tilde{I},\tilde{R})\in\mathcal{B}$. Then, it follows that

$$
\begin{aligned}
&\|F(X)-F(\tilde{X})\|_1\\
&=\left|f_1(X)-f_1(\tilde{X})\right|+\left|f_2(X)-f_2(\tilde{X})\right|+\left|f_3(X)-f_3(\tilde{X})\right|+\left|f_4(X)-f_4(\tilde{X})\right|\\
&=\left|\Lambda+\gamma R(t)-\left(\mu+\frac{\beta I(t)}{1+aI(t)}\right)S(t)-\right.\\
&\quad\left.\left(\Lambda+\gamma\tilde{R}(t)-\left(\mu+\frac{\beta\tilde{I}(t)}{1+a\tilde{I}(t)}\right)\tilde{S}(t)\right)\right|\\
&\quad+\left|\frac{\beta S(t)I(t)}{1+aI(t)}-(m+\rho)E(t)-\left(\frac{\beta\tilde{S}(t)\tilde{I}(t)}{1+a\tilde{I}(t)}-(m+\rho)\tilde{E}(t)\right)\right|\\
&\quad+\left|\rho E(t)-\left(\mu+\xi+\omega+\frac{\nu}{1+\eta I(t)}\right)I(t)\right.\\
&\quad\left.-\left(\rho\tilde{E}(t)-\left(\mu+\xi+\omega+\frac{\nu}{1+\eta\tilde{I}(t)}\right)\tilde{I}(t)\right)\right|\\
&\quad+\left|\left(\omega+\frac{\nu}{1+\eta I(t)}\right)I(t)-(\mu+\gamma)R(t)\right.\\
&\quad\left.-\left(\left(\omega+\frac{\nu}{1+\eta\tilde{I}(t)}\right)\tilde{I}(t)-(\mu+\gamma)\tilde{R}(t)\right)\right|\\
&\leq|\gamma|\left|R(t)-\tilde{R}(t)\right|+\left(|\mu|+\left|\frac{\beta}{a}\right|\right)\left|S(t)-\tilde{S}(t)\right|\\
&\quad+\left|\frac{\beta}{a}\right|\left|S(t)-\tilde{S}(t)\right|+(|m|+|\rho|)\left|E(t)-\tilde{E}(t)\right|\\
&\quad+|\rho|\left|E(t)-\tilde{E}(t)\right|+(|\mu|+|\xi|+|\omega|)\left|I(t)-\tilde{I}(t)\right|\\
&\quad+|\omega|\left|I(t)-\tilde{I}(t)\right|+(|\mu|+|\gamma|)\left|R(t)-\tilde{R}(t)\right|\\
&=\left(|\mu|+2|\gamma|\right)\left|R(t)-\tilde{R}(t)\right|+\left(|\mu|+2\left|\frac{\beta}{a}\right|\right)\left|S(t)-\tilde{S}(t)\right|\\
&\quad+\left(|m|+2|\rho|\right)\left|E(t)-\tilde{E}(t)\right|+\left(|\mu|+|\xi|+2|\omega|\right)\left|I(t)-\tilde{I}(t)\right|
\end{aligned}
$$

$$\leq \zeta\Big(\Big|R(t) - \tilde{R}(t)\Big| + \Big|S(t) - \tilde{S}(t)\Big| + \Big|E(t) - \tilde{E}(t)\Big| + \Big|I(t) - \tilde{I}(t)\Big| \Big)$$

$$= \zeta \left\| X - \tilde{X} \right\|_1,$$

where $\zeta = \max\left\{|\mu| + 2|\gamma|, \ \ |\mu| + 2\left|\frac{\beta}{a}\right|, \ \ |m| + 2|\rho|, \ \ |\mu| + |\xi| + 2|\omega|\right\}$. Thus F is locally Lipschitz and, therefore, by using Lemma 1 the dental caries infection model (1) admits a unique solution $\big(S(t), E(t), I(t), R(t)\big)$. □

Theorem 4. The solutions $S(t)$, $E(t)$, $I(t)$ and $R(t)$ of system (1) with initial values $S(0) = S_0 \geq 0$, $E(0) = E_0 \geq 0$, $I(0) = I_0 \geq 0$, $R(0) = R_0 \geq 0$ are positive for all $t > 0$.

Proof. The proof follows the same procedure as the one in Huo, Chen, and Wang [15]. □

F.1 Basic reproduction number

The basic reproduction number, R_0, is defined as the measure used to characterize the potential for an infectious organism to spread through a fully susceptible population, indicating the maximum potential rate of spread in that population. R_0 remains constant in the absence of prior immunity, while the effective reproduction number, R_t, decreases as immunity develops within the population. By employing the next-generation matrix technique [19], a basic reproduction number R_0 can be computed to obtain the rate of transmission of the model, which is given as

$$R_0 = \frac{\beta\rho\Lambda}{\mu(\mu+\rho)(\mu+\xi+\omega+\nu)}. \tag{16}$$

It stands for the typical quantity of secondary infections brought on by the original infection. We refer the reader to [19] for the basic reproduction number.

G Solution method and approximate results

In this section, we validate and demonstrate the applicability of the proposed strategy for solving a mathematical model of the dental caries infection in (1). All the numerical calculations are performed using Maple 2024 and the CPU time (in seconds) is provided for an Intel Core i7 3.00 GHz personal computer with 8 GB RAM. As the analytical solution for the given problem is unknown, we verify the reliability and accuracy of the technique by analyzing the residual error function. In order to draw the figures of approximations obtained for $S(t)$, $E(t)$, $I(t)$ and $R(t)$,

the following initial values and parameter values are considered [19]:

$$S_0 = 3,\ \ E_0 = 1,\ \ I_0 = 3,\ \ R_0 = 0,\ \ \Lambda = 0.70,\ \ \gamma = 0.000001,\ \ \mu = 0.13,$$
$$\beta = 0.05,\ \ a = 0.0002,\ \ m = 0.0077,\ \ \rho = 0.8,\ \ \xi = 0.05,\ \ \omega = 0.5,$$
$$\nu = 0.5,\ \ \eta = 0.000001.$$

The results obtained are depicted in Figures 3–6 with $m_1 = m_2 = m_3 = m_4 = 2$ (left side) and $m_1 = m_3 = 2$, $m_2 = m_4 = 3$ (right side) over a time period of 1 month. The runtime (in seconds) for our suggested technique and the optimal values for the residual function with several m_1, m_2, m_3 and m_4 values are provided in Tables 3 and 4. The mathematical model of dental caries infection (1) is also solved using the techniques discussed in section D, but with the Chebyshev polynomials neural network (CPNN). The behavior of the approximate solutions $S(t)$, $E(t)$, $I(t)$ and $R(t)$ with CPNN, $m_1 = m_2 = m_3 = m_4 = 2$ and $\beta = 0.07$ is shown in Figure 7. Table 4 lists the values obtained for the optimal residual function for various values of m_1, m_2, m_3 and m_4 using both CPNN and FNPNN over a time period of 1 month. The results presented in Figure 7 and Table 4 indicate that the proposed method is more accurate than the CPNN-based results.

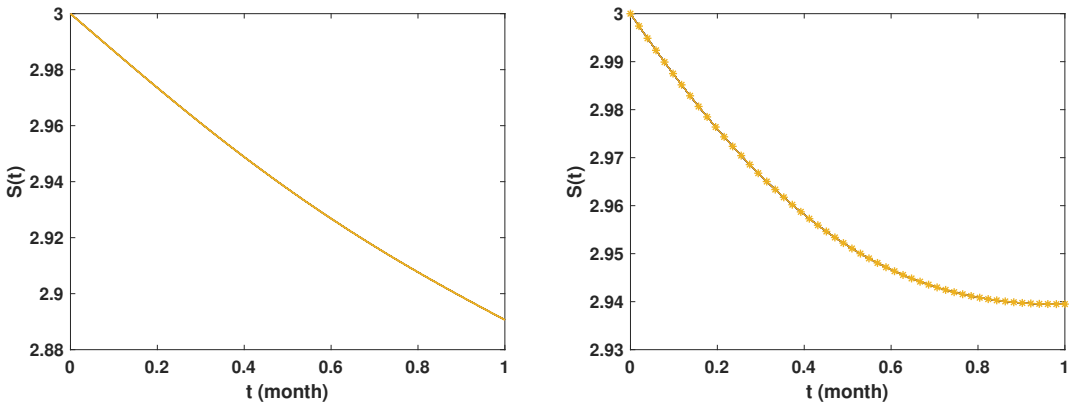


Figure 3: The FNPNN solutions of $S(t)$ with $m_1 = m_2 = m_3 = m_4 = 2$ (left side) and $m_1 = m_3 = 2$, $m_2 = m_4 = 3$ (right side).

Sensitivity analysis plays a central role in appraising the soundness and reliability of mathematical models across various disciplines. It helps distinguish which parameters have the most significant effect on outputs of the given model. By determining the sensitivity of the given model to different parameters, researchers can optimize the model structure or parameter values to achieve better performance. In this section, some scenarios will be designed to demonstrate

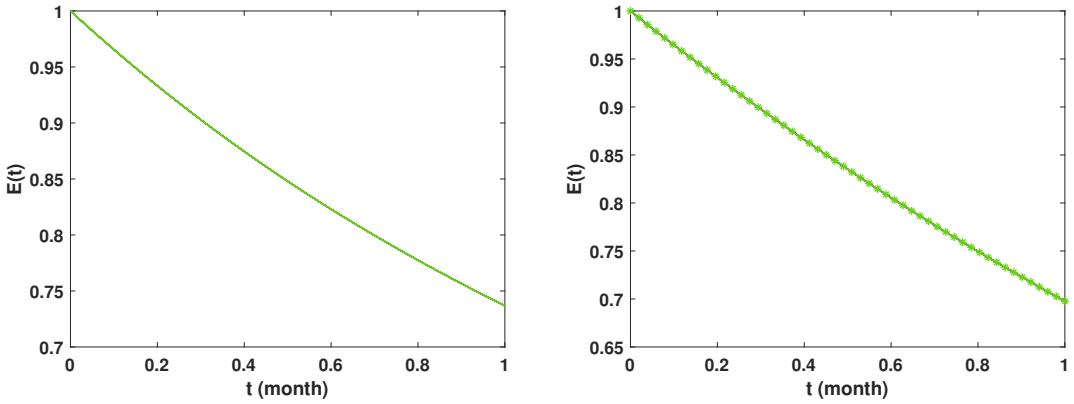


Figure 4: The FNPNN solutions of $E(t)$ with $m_1 = m_2 = m_3 = m_4 = 2$ (left side) and $m_1 = m_3 = 2$, $m_2 = m_4 = 3$ (right side).

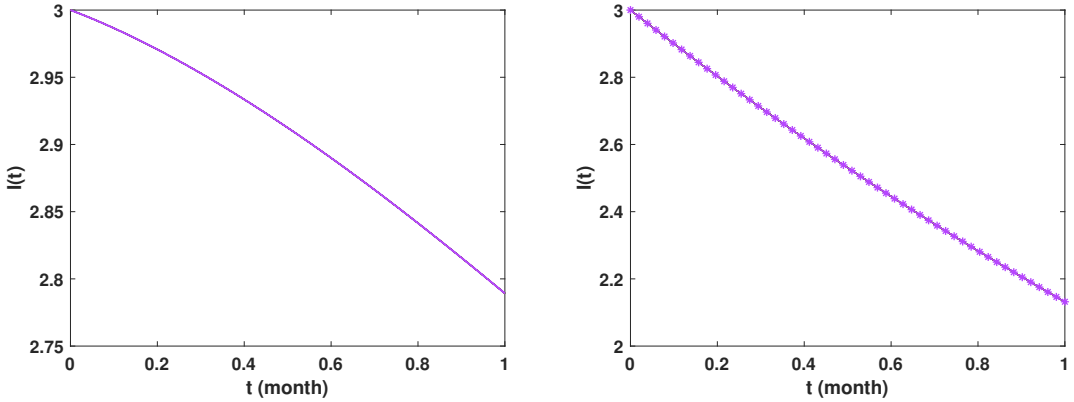


Figure 5: The FNPNN solutions of $I(t)$ with $m_1 = m_2 = m_3 = m_4 = 2$ (left side) and $m_1 = m_3 = 2$, $m_2 = m_4 = 3$ (right side).

Table 3: The runtime of the FNPNN method with several values of m_i, $i = 1, 2, 3, 4$.

case	m_1	m_2	m_3	m_4	CPU times
1	2	2	2	2	5.27
2	2	3	2	3	8.34
3	3	3	3	3	13.84

the sensitivity of the suggested model to diverse parameters.

1. Sensitivity to β: By choosing $\beta = 0.07$, model (1) is solved by the method stated in the previous subsection and figures of $S(t)$, $E(t)$, $I(t)$ and $R(t)$ are depicted in Figure 8 with $m_1 = m_2 = m_3 = m_4 = 2$ over a time period of 1 month.

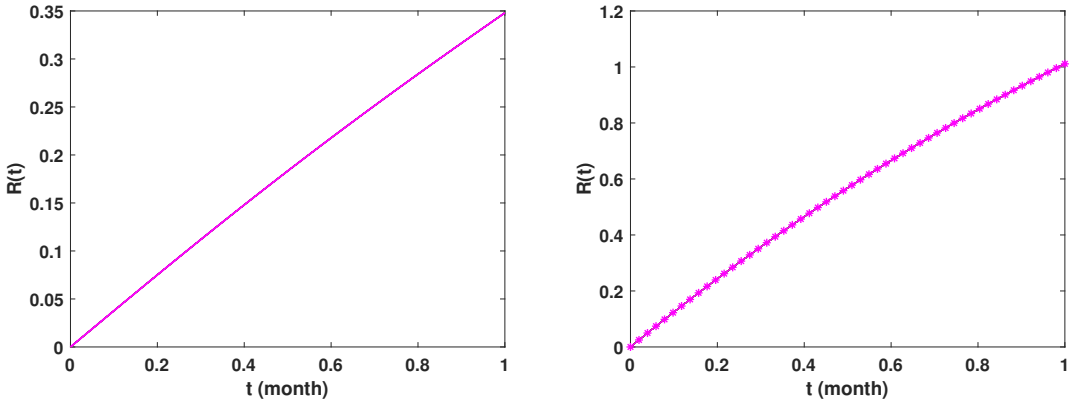


Figure 6: The FNPNN solutions of $R(t)$ with $m_1 = m_2 = m_3 = m_4 = 2$ (left side) and $m_1 = m_3 = 2$, $m_2 = m_4 = 3$ (right side).

Table 4: The optimal values of the residual function using CPNN and FNPNN with different values of m_i, where $i = 1, 2, 3, 4$.

case	m_1	m_2	m_3	m_4	Residual function (CPNN)	Residual function (FNPNN)
1	2	2	2	2	$5.8412E-03$	$2.0748E-07$
2	2	3	2	3	$2.8564E-04$	$3.2851E-08$
3	3	3	3	3	$9.5681E-05$	$3.7066E-10$

2. Sensitivity to Λ: By choosing values of Λ as $\Lambda = 0.40$, model (1) is solved by the method stated in the previous subsection and figures of $S(t)$, $E(t)$, $I(t)$ and $R(t)$ are depicted in Figure 9 with $m_1 = m_2 = m_3 = m_4 = 2$ over a time period of 1 month.

The number of susceptible human population $S(t)$ due to dental caries, the exposed human population but not infected $E(t)$ yet and the infected human population $I(t)$ decrease over time (Figures 3–5). The number of recovered human population through the treatment $R(t)$ increases (Figure 6). After applying the sensitivity analysis to β and Λ, we observe changes in Figures 8 and 9, indicating that model (1) is sensitive to these parameters.

Remark 3. For sensitivity analysis of the method, we refer the reader to [13].

H Concluding remarks

This study introduces a novel algorithm, FNPNN, designed to enhance the efficiency of optimization processes in solving the dental caries infection model. This study employs the optimization approach to demonstrate the feasibility of approximating the solution to the dental caries in-

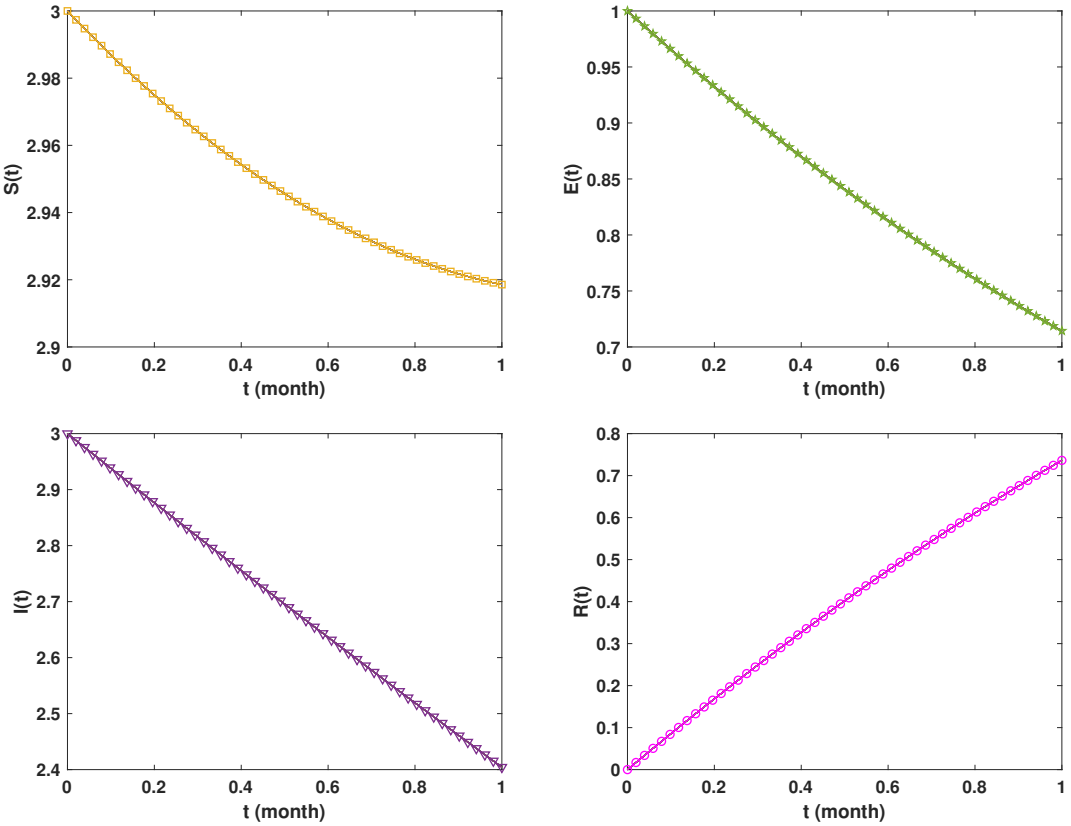


Figure 7: The CPNN solutions of $S(t)$, $E(t)$, $I(t)$ and $R(t)$ with $m_1 = m_2 = m_3 = m_4 = 2$ and $\beta = 0.07$.

fection model using an NN. The combination of the optimization method and NN refinement provides a promising framework for enhancing the precision and effectiveness of solving systems of differential equations in scientific and engineering applications. Tables 3 and 4 and Figures 3–9 show that the proposed model solution is reliable and produces the approximate solution with a minimal number of basis functions. In conclusion, numerical experiments are provided to deepen the understanding of the approach and to illustrate the detailed procedure for solving the dental caries infection model, thereby making the concepts clearer and more comprehensible.

Future research directions

The process of FNPNN by using the optimization method can be used to solve the fractional variable-order models, fractional optimal control problems, fractional-order partial differential

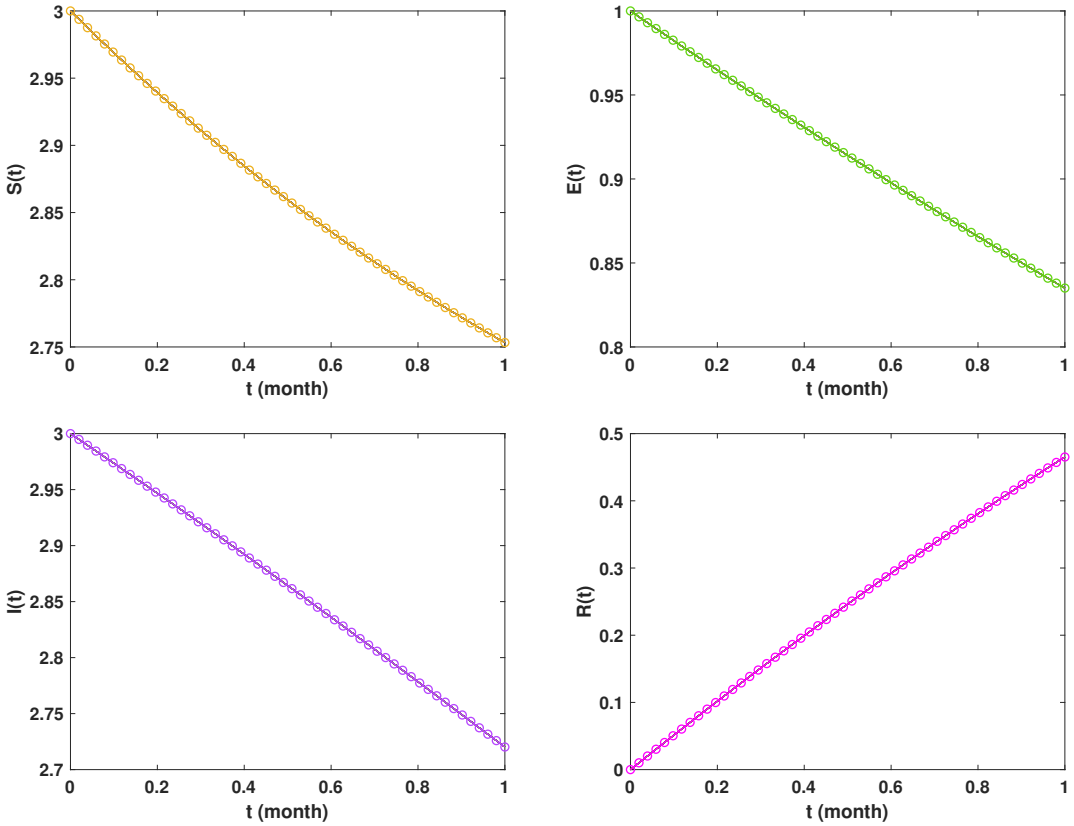


Figure 8: The FNPNN solutions of $S(t)$, $E(t)$, $I(t)$ and $R(t)$ with $m_1 = m_2 = m_3 = m_4 = 2$ and $\beta = 0.07$.

equations, fractional-order systems of partial differential equations, and various models in bioscience, etc.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflicts of interest.

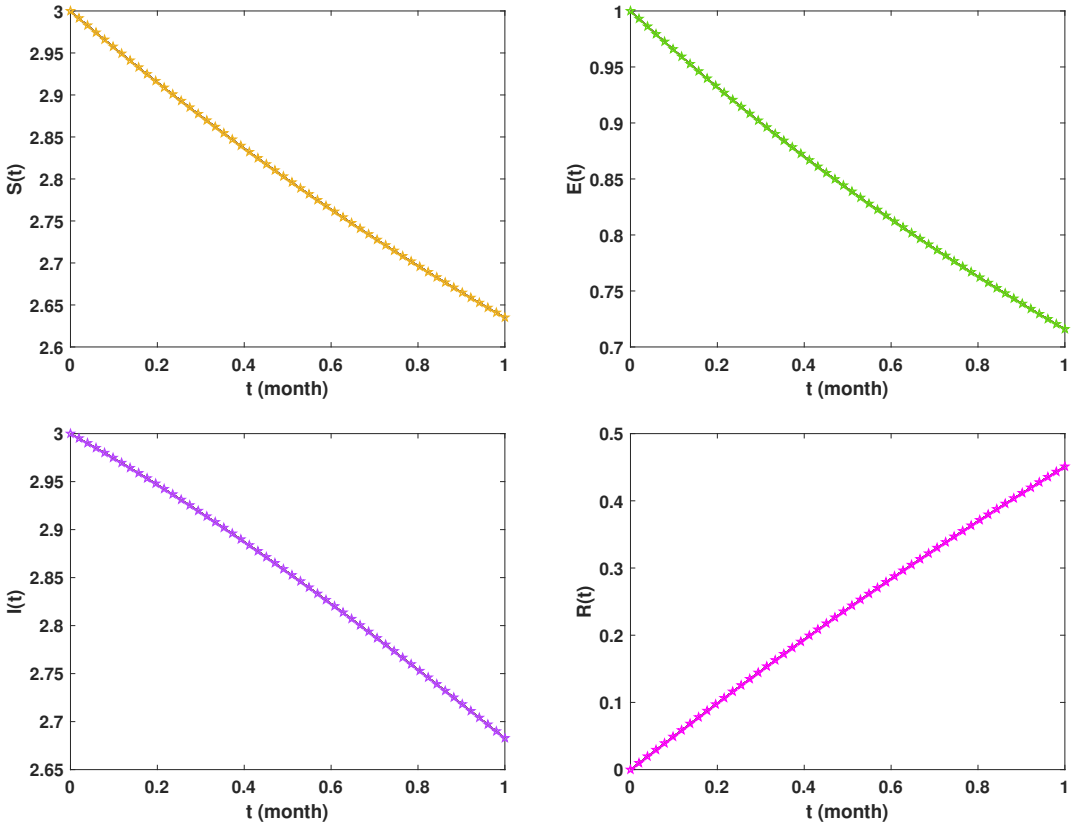


Figure 9: The FNPNN solutions of $S(t)$, $E(t)$, $I(t)$ and $R(t)$ with $m_1 = m_2 = m_3 = m_4 = 2$ and $\Lambda = 0.40$.

References

[1] Agapito, J., Mestre, A., Petrullo, P. and Torres, M.M., *Combinatorics of a generalized Narayana identity,* Linear Algebra Appl. 503 (2016) 56–82.

[2] Amit Prakash, R., *Numerical simulation of SIR childhood diseases model with fractional Adams–Bashforth method,* Math. Methods Appl. Sci. 46(12) (2023) 12340–12360.

[3] Anusha Bamini, A.M., Chitra, R., Brindha, D. and Chenthil Jegan, T.M. *A novel approach for premature detection of Alzheimer's disease using convolutional neural network in deep learning technique,* Wireless Pers. Commun. 135 (2024) 639–654.

[4] Avazzadeh, Z., Hassani, H., Agarwal, P., Mehrabi, S., Ebadi, M.J. and Dahaghin, M.S., *An optimization method for studying fractional-order tuberculosis disease model via generalized Laguerre polynomials,* Soft Comput. 27 (2023) 9519–9531.

[5] Avazzadeh, Z., Hassani, H., Ebadi, M.J., Zarei-Behjani, Z., Bayati Eshkaftaki, A., Mehrabi, S. and Haghighat, S., *Generalized Lerch polynomials: application in fractional model of CAR-T cells for T-cell leukemia,* Eur. Phys. J. Plus. 138(12) (2023) 1–13.

[6] Az-Zobi, E.A., Hussain, A.S., Iqbal, M., Aljohani, A., Tashtoush, M.A., Alsubaie, N.E., Al-Hamarneh, F.K.A., Nazar, N.M. and Baleanu, D., *Analytical and numerical solutions of MABC fractional advection dispersion models by utilizing the modified physics informed neural networks with impacts of fractional derivative,* Sci. Rep. 15 (2025) 43366.

[7] Az-Zobi, E.A., Shah, R., Alyousef, H.A., Tiofack, C.G.L. and El-Tantawy, S.A., *On the feed-forward neural network for analyzing pantograph equations,* AIP Adv. 14(2) (2024) 025042.

[8] Calin, O., *Deep learning architectures,* Springer, 2020.

[9] Chen, Z., Yang, Z. and Sheng, D., *Numerical analysis of linearly implicit Euler method for age-structured SIS model,* J. Appl. Math. Comput. 70 (2024) 969–996.

[10] Dawes, C. and Wong, D.T., *Role of saliva and salivary diagnostics in the advancement of oral health,* J. Dent. Res. 98(2) (2019) 133–141.

[11] Dhar, B. and Kumar Gupta, P., *A numerical approach of tumor-immune model with B cells and monoclonal antibody drug by multi-step differential transformation method,* Math. Methods Appl. Sci. 44(5) (2021) 4058–4070.

[12] Featherstone, J.D.B., *The caries balance: Contributing factors and early detection,* J. Calif. Dent. Assoc. 31(2) (2003) 129–133.

[13] Guo, Y., Ma, B. and Wu, R., *On sensitivity analysis of parameters for fractional differential equations with Caputo derivatives,* Electron. J. Qual. Theory Differ. Equ. 118 (2016) 1–17.

[14] He, J., *A new high-precision numerical method for solving the HIV infection model of CD4(+) cells*, Physica A. 653 (2024) 130090.

[15] Huo, H.F., Chen, R. and Wang, X.Y., *Modelling and stability of HIV/AIDS epidemic model with treatment,* Appl. Math. Modell. 40(13-14) (2016) 6550–6559.

[16] Iqbal, S., Qureshi, A.N., Alhussein, M., Aurangzeb, K., Zubair, M. and Hussain, A., *A novel reciprocal domain adaptation neural network for enhanced diagnosis of chronic kidney disease,* Expert Syst. 42(2) (2025) e13825.

[17] Kassebaum, N.J., Bernabe, E., Dahiya, M., Bhandari, B., Murray, C.J.L. and Marcenes, W., *Global burden of untreated caries: A systematic review and metaregression,* J. Dent. Res. 94(5) (2015) 650–658.

[18] Kassebaum, N.J., Smith, A.G.C., Bernabe, E., Fleming, T.D., Reynolds, A.E., Vos, T., Murray, C.J.L. and Marcenes, W., *Collaborators global, regional, and national prevalence, incidence, and disability-adjusted life years for oral conditions for 195 Countries, 1990-2015: A systematic analysis for the global burden of diseases, injuries, and risk factors,* J. Dent. Res. 96(4) (2017) 380–387.

[19] Khirsariya, S.R. and Chauhan, J.P., *Numerical and theoretical insights into fractional dental caries infection model,* Iran J. Sci. 49 (2025) 1605–1618.

[20] Kreyszig, E., *Introductory functional analysis with applications,* John Wiley and Sons. Inc. 1978.

[21] Li, Y. Chen, Y. and Podlubny, I., *Stability of fractional-order non linear dynamic systems: Lyapunov direct method and generalized Mittag-Leffler stability,* Comput. Math. Appl. 59(10) (2010) 1810–1821.

[22] Li, M. and Yang, Z., *Numerical analysis of an age-structured model for HIV viral dynamics with latently infected T cells based on collocation methods,* Math. Comput. Simul. 230 (2025) 289–305.

[23] Maji, C., Mukherjee, D. and Kesh, D., *Study of a fractional-order model of chronic wasting disease,* Math. Methods Appl. Sci. 43(7) (2020) 4669–4682.

[24] Mortazavi, S., Enshaei, Z. and Farajzadegan, Z., *Development of caries risk assessment tool for Iranian preschoolers: A primary validation study,* Int. J. Prev. Med. (2017) PMID :29184643.

[25] Peres, M.A., Macpherson, L.M.D., Weyant, R.J., Daly, B., Venturelli, R., Mathur, M.R., Listl, S., Celeste, R.K., Guarnizo-Herreno, C.C., Kearns, C., Benzian, H., Allison, P. and Watt, R.G., *Oral diseases: A global public health challenge,* Lancet. 394(10194) (2019) 249–260.

[26] Petkovic, M.D., Barry, P. and Rajkovic, P., *Closed-form expression for Hankel determinants of the Narayana polynomials,* Czech. Math. J. 62(1) (2012) 39–57.

[27] Pitts, N.B., Zero, D.T., Marsh, P.D., Ekstrand, K., Weintraub, J.A., Ramos-Gomez, F., Tagami, J., Twetman, S., Tsakos, G. and Ismail, A., *Dental caries,* Nat. Rev. Dis. Primers. 3 (2017) 17030.

[28] Priyanga, P., Pattankar, V.V. and Sridevi, S., *A hybrid recurrent neural network-logistic chaos-based whale optimization framework for heart disease prediction with electronic health records,* Comput. Intell. 37(1) (2021) 315–343.

[29] Rafique, M., Ur Rehamn, M.A., Rafiq, M., Iqbal, Z., Ahmed, N., Alhazmi, H., Niazai, S. and Khan, I., *Time delayed fractional diabetes mellitus model and consistent numerical algorithm,* Sci. Rep. 14 (2024) 23871.

[30] Rahimkhani, P. and Ordokhani, Y., *Orthonormal Bernoulli wavelets neural network method and its application in astrophysics,* Comput. Appl. Math. 40 (2021) 78.

[31] Ruttanaprommarin, N., Sabir, Z., Sandoval Nunez, R.A., Az-Zobi, E.A., Weera, W., Botmart, T. and Zamart, C., *A stochastic framework for solving the prey-predator delay differential model of Holling type-III,* Comput. Mater. Continua. 74(3) (2023) 5915–5930.

[32] Sabir, Z., Athar Mehmood, M., Umar, M., Salahshour, S., Altun, Y., Arbi, A. and Ali, M.R., *A numerical treatment through Bayesian regularization neural network for the chickenpox disease model,* Comput. Biol. Med. 187 (2025) 109807.

[33] Sabir, Z., Ben Said, S. and Al-Mdallal, Q., *A fractional order numerical study for the influenza disease mathematical model,* Alexandria Engineering Journal, 65 (2023) 615–626.

[34] Shafik, W., Tufail, A., Silva, C.L.D. and Haji Mohd Apong, R.A.A., *A novel hybrid inception-xception convolutional neural network for efficient plant disease classification and detection,* Sci. Rep. 15 (2025) 3936.

[35] Sheiham, A., *Dietary effects on dental diseases,* Public Health Nutr. 4(2B) (2001) 569–591.

[36] Sheiham, A. and James, W.P.T., *A new understanding of the relationship between sugars, dental caries and fluoride use: implications for limits on sugars consumption,* Public Health Nutr. 17(10) (2014) 2176–2184.

[37] Slobodnikova, L., Halasa, A.I., Kalmanova, S., Calkovsky, B., Juricek, R., Malachovsky, I., Repiska, V., Skerenova, M. and Janickov, M., *Genetic factors influencing susceptibility to dental caries*, Bratisl. Med. J. 125 (2024) 635–647.

[38] Verma, A. and Kumar, M., *Numerical solution of Bagley-Torvik equations using Legendre artificial neural network method,* Evol. Intell. 14 (2021) 2027–2037.

[39] Verma, A. and Kumar, M. *Numerical solution of third-order Emden–Fowler type equations using artificial neural network technique,* Eur. Phys. J. Plus. 135(9) (2020) 751.

[40] Verma, A., Sumelka, W. and Yadav, P.K. *The numerical solution of nonlinear fractional Lienard and Duffing equations using orthogonal perceptron,* Symmetry, 15(9) (2023) 1753.

[41] Wang, X., Willing, M.C., Marazita, M.L., Wendell, S., Warren, J.J., Broffitt, B., Smith, B., Busch, T., Lidral, A.C. and Levy, S.M., *Genetic and environmental factors associated with dental caries in children: The Iowa Fluoride Study,* Caries Res. 46(3) (2012) 177–184.

[42] Wang, Z., Yang, Z., Yang, G. and Zhang, C., *Numerical analysis of age-structured HIV model with general transmission mechanism,* Commun. Nonlinear Sci. Numer. Simul. 134 (2024) 108020.

[43] Wang, X., Zhang. P., Lu, H., Luo, D., Yang, D., Li, K., Qiu, S., Zeng, H. and Zeng, X., *Risk prediction models for dental caries in children and adolescents: a systematic review and meta-analysis,* BMJ Open. 15(3) (2025) e088253.

Aims and scope

Iranian Journal of Numerical Analysis and Optimization (IJNAO) is published by the Department of Applied Mathematics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad. Papers dealing with different aspects of numerical analysis and optimization, theories and their applications in engineering and industry are considered for publication.

Journal Policy

All submissions to IJNAO are first evaluated by the journal's Editor-in-Chief or one of the journal's Associate Editors for their appropriateness to the scope and objectives of IJNAO. If deemed appropriate, the paper is sent out for review using a single blind process. Manuscripts are reviewed simultaneously by reviewers who are experts in their respective fields. The first review of every manuscript is performed by at least two anonymous referees. Upon the receipt of the referee's reports, the paper is accepted, rejected, or sent back to the author(s) for revision. Revised papers are assigned to an Associate Editor who makes an evaluation of the acceptability of the revision. Based upon the Associate Editor's evaluation, the paper is accepted, rejected, or returned to the author(s) for another revision. The second revision is then evaluated by the Editor-in-Chief, possibly in consultation with the Associate Editor who handled the original paper and the first revision, for a usually final resolution.

The authors can track their submissions and the process of peer review via: http://ijnao.um.ac.ir

All manuscripts submitted to IJNAO are tracked by using "iThenticate" for possible plagiarism before acceptance.

Instruction for Authors

The Journal publishes all papers in the fields of numerical analysis and optimization. Articles must be written in English.
All submitted papers will be refereed and the authors may be asked to revise their manuscripts according to the referee's reports. The Editorial Board of the Journal keeps the right to accept or reject the papers for publication.
The papers with more than one authors, should determine the corresponding author. The e-mail address of the corresponding author must appear at the end of the manuscript or as a footnote of the first page.

It is strongly recommended to set up the manuscript by Latex or Tex, using the template provided in the web site of the Journal. Manuscripts should be typed double-spaced with wide margins to provide enough room for editorial remarks.

References should be arranged in alphabetical order by the surname of the first author as examples below:

[1] Brunner, H. *A survey of recent advances in the numerical treatment of Volterra integral and integro-differential equations*, J. Comput. Appl. Math. 8 (1982), 213-229.

[2] Stoer, J. and Bulirsch, R. *Introduction to Numerical Analysis*, Springer-verlag, New York, 2002.

Iranian Journal of Numerical Analysis and Optimization

CONTENTS Vol. 17, No. 1, pp 1-628

Adaptive hybrid numerical schemes with singularity treatment for third order Lane–Emden-type equations 1
M. Fazli

Unraveling the relationship between COVID-19 viral load, disease severity, and innate immune response: Insights from mathematical modeling and sensitivity analysis 25
M. Lotfi, A. Jabbari, R. Rahbarghazi and H. Kheiri

A stable and high-order accurate D-RBF-PU method via hybrid kernels 45
M. Esmaeilbeigi and R. Cavoretto

Behavioral modeling and control strategies for Rumor-induced customer dynamics on digital platforms 72
S. Belhdid, H. Hassouni, O. Aarabate and O. Balatif

Understanding the impact of vaccination on Omicron variants 109
G.P. Sahu, A.S. Thakur, C.S. Sikarwar and H. Singh

A modified school bullying model with a recovery compartment: Stability analysis and optimal control 145
A. Sasmito, C. Miftahurrizki K H and R. Maeni

Modeling the transmission dynamics of meningococcal meningitis under hospitalization effects 177
I.O. Joshua, K. Kiruthika, M. Gümüş, A. Raza, E. Chandrakumar and N. Jeeva

A predator-prey model incorporating fear effect, prey vigilance costs, and predator hunting cooperation: Analysis and simulations 217
A. Akinsanya

Algebraic properties of the new polyconvolution operator related to the Laplace, Fourier cosine integral transforms and applications 247
M.K. Nguyen, L.T.P. Bui and V.T. Tran

Mathematical modeling and optimal control of hepatitis B virus dynamics . . 273
L. El Houari, M. Elkaf and J. Danane

Memory-driven fractional dynamics of pneumonia transmission with prophylactic intervention 290
K. Soni, S. Kumar, S. Kumawat and A.K. Sinha

A conjugate gradient algorithm based on a combined descent scheme for two-layer neural network training and unconstrained optimization 332
C. Tebbal, M. Laouar, R. Ziadi and A.E. Mehamdia

Enhancing conjugate gradient performance via a diagonal quasi-Newton strategy: Applications in adaptive filtering and power allocation 359

N. Bidabadi

Mathematical analysis of Mpox transmission using an SAIR-SI model with optimal control strategies 380
T. Janan, F. Fatmawati and A. Hasan

Stability analyses of block hybrid methods for numerical solutions of differential equations 416
O. Adedire and U.W. Sirisena

An approach to find the optimal design of a 3D damping set for a damped wave equation in spherical coordinates 457
H. Alimorad, A. Fakharzadeh Jahromi and M. Goodarzi

Stability and sensitivity analysis of a nonlinear compartmental model for banking risk contagion 494
S. Ibnoukady, A. Meskaf and K. Allali

Mathematical modeling, global stability, sensitivity analysis, and optimal control of merchant adoption dynamics in e-commerce 529
H. Hassouni, A. El Bhih, O. Aarabate and O. Balatif

Global dynamics of a fractional-order Hantavirus model with awareness-driven protective behavior 570
Z. Yaagoub

An appropriate fractional Narayana polynomials neural network method for dental caries infection model 603
H. Hassani, Z. Avazzadeh, A. Turan-Dincel and H. Farahmand

web site: https://ijnao.um.ac.ir
Email: ijnao@um.ac.ir
ISSN-Print: 2423-6977
ISSN-Online: 2423-6969