frontiers | Research Topics

Advances in data-driven approaches and modeling of complex systems

Edited by
Mohd Hafiz Mohd, Joel M. Addawe, Hirohide Haga, Junpyo Park and Tri Nguyen-Huu

Published in
Frontiers in Applied Mathematics and Statistics

ISSN 1664-8714
ISBN 978-2-8325-2665-1
DOI 10.3389/978-2-8325-2665-1

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Advances in data-driven approaches and modeling of complex systems

Topic editors

Mohd Hafiz Mohd — Universiti Sains Malaysia, Malaysia
Joel M. Addawe — University of the Philippines Baguio, Philippines
Hirohide Haga — Doshisha University, Japan
Junpyo Park — Kyung Hee University, Republic of Korea
Tri Nguyen-Huu — Institut de Recherche Pour le Développement (IRD), France

Citation

Mohd, M. H., Addawe, J. M., Haga, H., Park, J., Nguyen-Huu, T., eds. (2023). *Advances in data-driven approaches and modeling of complex systems*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-2665-1

Table of contents

frontiers | Frontiers in Applied Mathematics and Statistics

TYPE Editorial
PUBLISHED 25 May 2023
DOI 10.3389/fams.2023.1215077

OPEN ACCESS

EDITED AND REVIEWED BY
Daniel Potts,
Chemnitz University of Technology, Germany

*CORRESPONDENCE
Mohd Hafiz Mohd
✉ mohdhafizmohd@usm.my

RECEIVED 01 May 2023
ACCEPTED 11 May 2023
PUBLISHED 25 May 2023

CITATION
Mohd MH, Nguyen-Huu T, Park J, Addawe JM and Haga H (2023) Editorial: Advances in data-driven approaches and modeling of complex systems.
Front. Appl. Math. Stat. 9:1215077.
doi: 10.3389/fams.2023.1215077

Editorial: Advances in data-driven approaches and modeling of complex systems

Mohd Hafiz Mohd[1]*, Tri Nguyen-Huu[2], Junpyo Park[3], Joel M. Addawe[4] and Hirohide Haga[5]

[1]School of Mathematical Sciences, Universiti Sains Malaysia, Penang, Malaysia, [2]UMMISCO, Sorbonne Université, Institut de Recherche Pour le Développement, IRD, Bondy, France, [3]Department of Applied Mathematics, College of Applied Science, Kyung Hee University, Yongin, Republic of Korea, [4]Department of Mathematics and Computer Science, University of the Philippines Baguio, Baguio, Philippines, [5]Faculty of Science and Engineering, Doshisha University, Kyoto, Japan

KEYWORDS

complex systems analysis, modeling and simulation, deterministic model, stochastic model, machine learning, big data analytics, mathematical model, persistent homology

Editorial on the Research Topic
Advances in data-driven approaches and modeling of complex systems

This Research Topic discusses the latest developments in complex systems research and intends to give exposure to prospective readers about the data-driven approaches and modeling aspects of distinct complex physical, biological, and social systems. To address these issues, different modeling techniques e.g., ordinary and partial differential equations, discrete maps and stochastic models are employed to examine the complex systems of interest. Additionally, data-driven techniques e.g., agent-based modeling, machine learning, persistent homology, and big data analytics have also been discussed in greater detail so as to provide another paradigm on how data analysis and data interpretation are being performed in complex systems research. Special attention is given to various applications on the techniques of complex systems in examining the socio-environmental issues, epidemiology, ecological and physical problems, engineering, agriculture and financial applications. Many realistic examples from recent research are also employed in this Research Topic as illustrations. The main purpose of this article collection is to emphasize a unified approach to complex systems analysis, which goes beyond examining complicated phenomena of numerous real-life systems; this is done by investigating a huge number of components that interact with each other at different microscopic and macroscopic scales; new insights and emergent collective behaviors can evolve from the interactions between individual components and also with their environments. These tools and concepts will allow us to better understand the patterns of various real-life systems and help us comprehend the mechanisms behind which distinct factors shape some complex systems phenomena. As mentioned above, this Research Topic is specially designed to take into account a multidisciplinary approach in complex systems analysis that will encourage the transfer of ideas and methodology from data-driven techniques and modeling fields to the other areas of knowledge (and vice versa).

One of the main themes received by this Research Topic is on the modeling of COVID-19 epidemiology and the complex nature of diseases' progression. In general, the simplest mathematical frameworks that are widely used in infectious disease modeling is the deterministic ordinary differential equations (ODE) model. Traditional ODE models of disease transmission are often employed to predict disease trajectories and evaluate

the effectiveness of alternative treatment or intervention strategies. For instance, in the recent COVID-19 global outbreaks, differential equations-based model constitutes an important tool to examine the effectiveness of (non-)pharmaceutical interventions and to guide policymaking in different countries. In this line of research, Aini et al. employed compartmental epidemic ODE to examine the effects of mass mobility on the COVID-19 transmission between localities in Indonesia. One of the contributions of this this study to complex system modeling is to formulate a useful method of estimating the movement of people by representing mobility in the form of a function; it is discovered that the predictions obtained using this technique yield a reasonably good result, compared with the actual data. Another modeling work by Inayaturohmat et al. also used compartmental modeling to investigate the combined influences of co-infection with tuberculosis and distinct intervention measures on the COVID-19 outbreaks. The optimal control approach has been incorporated into this modeling framework to study the effectiveness of isolation and treatment on the prevention for COVID-19 and tuberculosis. Other possible determinants of COVID-19 disease transmission are the impacts of stochasticity, vaccination strategies and also the reinfection problem. These factors have been considered by Campos et al. where they have developed stochastic differential equations model to envisage the interplay of stochasticity, booster shots and waning of immunity on the control of COVID-19. Their results show the importance of booster shots that enhance the vaccine-induced immunity duration. Incorporating uncertainties in the COVID-19 simulation can be beneficial, particularly for the policy makers to obtain better projection in designing more robust and refined intervention strategies.

Apart from incorporating stochasticity into differential equation-based model, other realisms can also be included to gain realistic insights on the physical and biological phenomena of interest. Inclusion of spatial diffusion component through the use of partial differential equations (PDE) model has played a pivotal role in better understanding the influences of spatial and dispersion process in shaping dynamical behavior of certain complex systems. For instance, Babajanov and Abdikarimov employed a PDE system to investigate the dynamics of traveling wave solutions using the functional variable method. The proposed method is found to be effective in constructing the exact traveling wave solutions of non-linear wave equations arising in mathematical physics and engineering. Additionally, another realism such as the growth process can also be considered in formulating the deterministic systems, and this approach has been widely used in the biological applications. In general, the growth of species can happen either by continuous breeding or this process can also occur seasonally at discrete times. Examples of species that breed seasonally are annual plants and insects; and apes are an example of species that breed continuously. Thus, depending on the type of growth process, this consideration can lead to distinct deterministic systems: (i) continuous-time models (e.g., ODE and PDE, for continuous breeding process), which have been discussed thoroughly in the previous paragraph; (ii) discrete-time models (e.g., discrete maps, for seasonal breeding process). For the contribution on discrete-time modeling framework, Hasibuan et al. have examined the competitive outcomes of some semelparous species and they discovered that the local stability condition of the coexistence equilibrium is determined by the degree of interspecific and intraspecific competition. This intriguing observation may have some applications in comprehending the dynamics of complex natural ecosystems and also crucial in devising sustainable management plans of the natural resources, such as in the fisheries industries.

Most of the aforementioned modeling studies discussed are based on the mechanistic system paradigms, meaning that these models are being constructed using the fundamental laws of natural sciences, including physical and biological principles. From practical viewpoints, these models can be employed to gain some beneficial insights on our complex natural world, particularly in the case of limited real data. However, the availability of vast and ever-increasing quantities of data has invoked a new paradigm in modeling and simulation fields with the advancement of data-driven modeling techniques such as agent-based modeling (ABM), machine learning, persistent homology, and big data analytics. In this area of research, Vinyals et al. proposed an integrated modeling approach that combined the ABM and machine learning techniques with related applications in digital technologies for agriculture. Another contribution on the agent-based modeling reviewed the state-of-the-art techniques on the inclusion of multiple levels of analysis, abstraction, and representation in the ABM frameworks. This innovative technique is often helpful in order to analyze different physical domains and applications (e.g., urban planning, land-use change, adaptation to environmental changes, biodiversity protection in socio-ecosystems, environmental pollution control, etc.) where interactions between individuals and their environments can give rise to emergent phenomena that are difficult to study otherwise. Another kind of data-driven modeling approach is the persistent homology (PH), which extracts the topological features from non-linear time series data. This technique is particularly useful for deeper understanding of the dynamical systems behavior and early detection or warning signals of catastrophe. Motivated by the capability of PH, Ismail et al. utilized this method combined with critical slowing down and some statistical correlation tests to examine the early warning signals of financial crises. Overall, PH approach provides a robust and accurate analysis of a dynamical system as this technique reveals the system topology at different spatial resolutions. This often leads to the persistence of only the major features of the underlying system and removes spurious effects of noise. It can also be seen that the big data analytics and statistical techniques play significant roles in the advent of data-driven modeling frameworks of distinct complex systems. This theme has inspired the formulation of a hybrid modeling system of autoregressive integrated moving average and fuzzy time series Markov chains by Devianto et al. to study the long pattern of the crude oil movement price using data from the West Texas Intermediate oil price for the year 2003 till 2021.

The modeling frameworks and problems presented in this Research Topic offer an overall outlook of the many facets of modeling approaches, analysis, and practical implications. However, the works covered in this Research Topic are far from exhaustive. Rather, they point to a selected range of problems that can be addressed by the mathematical modeling and data-driven

computational approaches, and hopefully this insight can inspire the development of novel tools and techniques in the future.

Author contributions

MM, TN-H, JP, JA, and HH conceived and wrote the editorial. All authors contributed to the article and approved the submitted version.

Funding

MM acknowledges the support from the Fundamental Research Grant Scheme with Project Code: FRGS/1/2022/STG06/USM/02/1 by the Ministry of Higher Education, Malaysia (MOHE).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

frontiers | Frontiers in Applied Mathematics and Statistics

ORIGINAL RESEARCH
published: 03 June 2022
doi: 10.3389/fams.2022.912150

The Impact of Mass Exodus on the Resurgence of COVID-19 Cases: Case Study of Regions in Indonesia

Wirdatul Aini[1], *Kamal Khairudin Sukandar*[1]*, *Nuning Nuraini*[1,2] *and Dewi Handayani*[1,2]

[1] Mathematics Department, Faculty of Mathematics and Natural Sciences, Institut Teknologi Bandung, Bandung, Indonesia, [2] Center of Mathematical Modeling and Simulation, Institut Teknologi Bandung, Bandung, Indonesia

Consideration of human mobility is essential for understanding the behavior of COVID-19 spread, especially when millions of people travel across borders around Eid al-Fitr. This study aims to grasp the effect of mass exodus between regions on active cases of COVID-19 through a mathematical perspective. We constructed a multiregional SIQRD (susceptible–infected–quarantined–recovered–death) model that accommodates the direct transfer of people from one region to others. The mobility rate was estimated using the proposed Dawson-like function, which requires data from an origin–destination matrix. Assuming that only susceptible, inapparently infected, and recovered individuals travel around Eid al-Fitr, the rendered model well-depicted the actual data at that time, giving either a significant spike or decline in the number of active cases due to the mass exodus. Most agglomerated regions such as Jakarta and Depok City experienced a fall in active case numbers, both in actual data and in the simulated model. However, most rural areas experienced the opposite, such as Bandung District and Cimahi City. This study confirmed that most travelers journeyed from big cities to the rural regions, and it scientifically demonstrated that mass mobility affects COVID-19 transmission between areas.

Keywords: COVID-19, mobility aspect, Eid al-Fitr, SIQRD model, OD matrix, Dawson integral function

OPEN ACCESS

Edited by:
Joel M. Addawe, University of the Philippines Baguio, Philippines

Reviewed by:
Khalid Hattaf, Centre Régional des Métiers de l'Education et de la Formation (CRMEF), Morocco
Haitao Song, Shanxi University, China

***Correspondence:**
Kamal Khairudin Sukandar
kamalkhairudin@students.itb.ac.id

Specialty section:
This article was submitted to Mathematics of Computation and Data Science, a section of the journal Frontiers in Applied Mathematics and Statistics

Received: *04 April 2022*
Accepted: *16 May 2022*
Published: *03 June 2022*

Citation:
Aini W, Sukandar KK, Nuraini N and Handayani D (2022) The Impact of Mass Exodus on the Resurgence of COVID-19 Cases: Case Study of Regions in Indonesia. Front. Appl. Math. Stat. 8:912150. doi: 10.3389/fams.2022.912150

INTRODUCTION

The COVID-19 pandemic at the beginning of 2020 echoed across all countries. That high-spreader virus has led leaders in almost all countries to apply similar policies: restricting people's movement except for grocery shopping. This policy is expected to reduce human-to-human interactions, leading to the decline of COVID-19 case numbers [1]. This restriction lasted for most of 2020, even though some relaxation had occurred in some countries. In India, the government eased the lockdown after a significant decline in COVID-19 cases, allowing more people to be mobile. A surge in COVID-19 cases was identified after the first relaxation, which was believed to be a direct implication of the relaxation [2].

The Centers for Disease Control and Prevention (CDC) has been consistently providing recommendations regarding restrictions on the movement of people to suppress the spread of the virus in the population since March 8th, 2021 [3]. These restrictions on movement are based on several factors: the number of COVID-19 cases in the community, exposure during travel, the level of crowding in a place, and so on. It is clear that the higher the number of COVID-19 cases in a community and the more crowded a place is, the greater the chance of being exposed to the virus. This provides a strong rationale for restricting movement to suppress the spread of infection.

In line with what the CDC recommends, movement restrictions have also been implemented in several provinces of Indonesia with high numbers of COVID-19 cases. For example, Jakarta employs PSBB (large-scale social restrictions), which limits the movement of Jakarta residents through restrictions on schools and workplaces, cessation of religious activities in places of worship, restrictions on activities in public places or facilities, restrictions on sociocultural activities, and on transportation activities. By the end of 2020, the PSBB had been implemented 12 times in Jakarta [4]. By the end of February 2021, the macro and micro PPKM (enforcement of community activity restrictions) was implemented to respond to a significant increase in cases in Jakarta, reaching more than 3,000 new cases per day [5]. In general, PPKM is similar to PSBB but applies to the Java and Bali regions.

Ahead of Eid al-Fitr 1442 Hijri, a mass exodus (a trip to one's hometown) is unavoidable, which will lead millions of people to leave on their odysseys [6]. Even though there was a significant drop in the number of travelers in 2020, it was expected that the number would rebound in 2021 [7]. The mass mobility was believed to trigger more infections since the inapparently infected could travel across borders. As shown earlier in China, the human mobility due to Chinese New Year did result in variations in disease incidence between cities. Cities that were closer to the epidemic center and with a higher population in urban areas faced higher risks of disease incidence [8].

In this study, a multiregional mathematical model that accommodates the mobility of people between regions was constructed. The model was a generalized SIQRD model that split the population into five groups: susceptible, infected, quarantined, recovered, and deaths. The additional parameters representing the time-dependent mobility rate from one region to another were depicted by the modified Dawson function. Such parameters were estimated by means of the proposed algorithm requiring data from an origin–destination matrix. The numerical simulations of the constructed model depicted how the mobility rate between regions could affect the surge or decline in the number of COVID-19 cases.

CONTEXT

Mudik: Indonesia's Annual Mass Exodus

Mudik or Balik Kampung are common terms used by Indonesian people to represent migrants or migrant workers returning to their hometowns or villages before significant holidays, such as Lebaran (Eid al-Fitr) or Chinese New Year [9]. The primary motivation of this tradition is to visit family, especially parents. Further, many of them return to their homes to attend a rare opportunity: a gathering of members of their extended families who are seldomly seen and are typically scattered in other cities, other provinces, or even overseas. *Mudik* Eid al-Fitr or similar traditions can also be found in other Muslim-majority countries, such as Egypt and Bangladesh [10].

The average number of total travelers in such exoduses in Indonesia, especially during Eid al-Fitr, is ∼30 million [11].

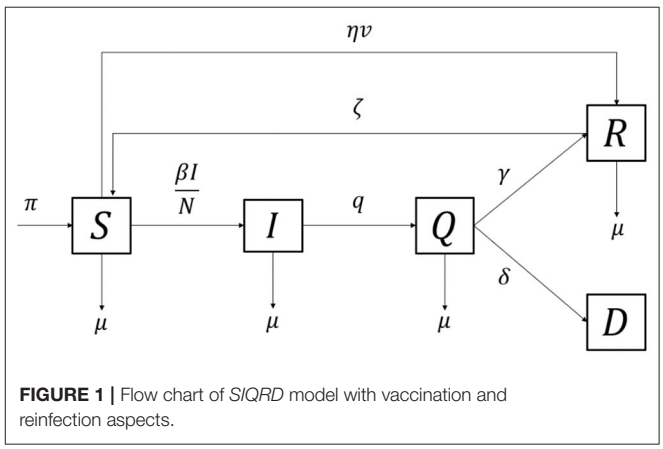


FIGURE 1 | Flow chart of *SIQRD* model with vaccination and reinfection aspects.

TABLE 1 | Descriptions of variables and parameters in the $\{SIQRD\}_k$ model.

Notation	Description	Unit	Value
$S_k(t)$	Number of susceptible individuals at time t in region k	Individual	-
$I_k(t)$	Number of infected, yet not quarantined, individuals at time t in region k	Individual	-
$Q_k(t)$	Number of quarantined individuals at time t and region k	Individual	-
$R_k(t)$	Number of recovered individuals after quarantine for COVID-19 infection at time t and region k	Individual	-
$D_k(t)$	Number of deceased individuals after quarantine for COVID-19 infection at time t and region k	Individual	-
β_k	Transmission rate of COVID-19 in region k	1/day	Estimated
γ_k	Recovery rate for quarantined individuals in region k	1/day	Estimated
δ_k	Death rate for quarantined individuals in region k	1/day	Estimated
$\alpha_{jk}(t)$	Daily mobility rate from region j to k at time t due to the mass exodus	1/day	Estimated
v_k	Daily average of rate of inoculation conducted in region k	1/day	Estimated
ζ	Reinfection rate of COVID-19 due to natural immunity waning	1/day	Assumed
q	Quarantine rate	1/day	Assumed
η	Vaccine efficacy	-	Assumed
π	Natural recruitment rate	1/day	Assumed
μ	Natural death rate	1/day	Assumed

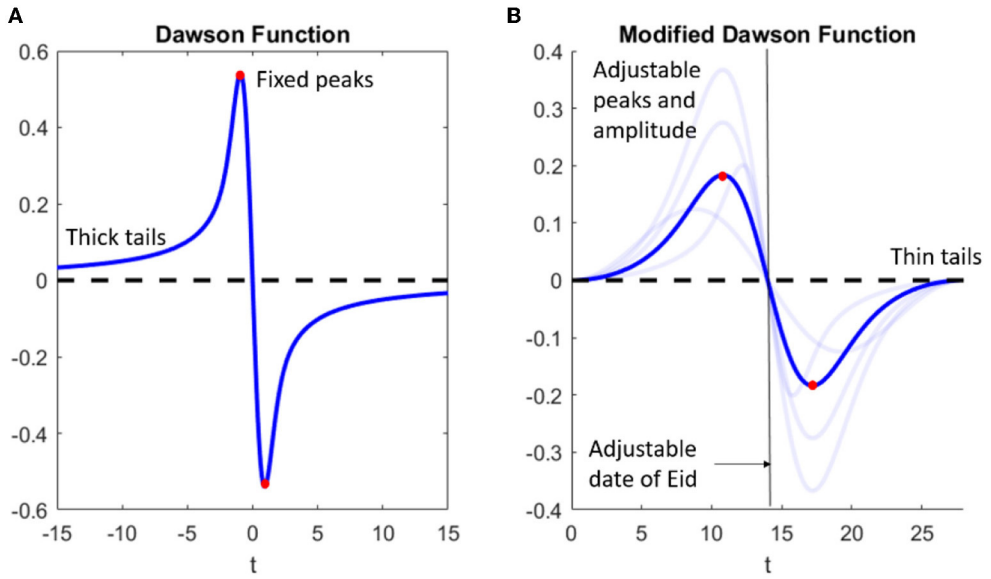


FIGURE 2 | Comparison of the generic Dawson Integral function **(A)** with its modification **(B)**.

Even though the travel related to the *mudik* homecoming before Lebaran takes place in most Indonesian urban centers, the emphasis is on the nation's largest urban agglomeration Greater Jakarta, as millions of Jakartans exit the city by various means of transportation, overwhelming train stations and airports, and also clogging highways. Annually, the number of travelers during the exodus shows a constant pattern: gradually increasing from 2 weeks prior to the holy day and reaching its spike 2 days before it. The reversed mobility pattern is also shown days after the holiday, representing the return of migrants from their destination [12].

Every year, the total number of Eid al-Fitr homecomings has shown a significant increase except in 2020 [13], as the number of travelers dropped to almost zero, given the global pandemic. However, based on the survey data obtained by Angkutan Lebaran Research Group, the expected number of travelers on the 2021 Eid al-Fitr *mudik* will have increased and thus could have increased the chance of infection from COVID-19 due to the increasing rate of person-to-person interactions. Even worse, Jakarta was projected to contain more than 50% of the total recorded cases in Indonesia by April 2021; consequently, the *mudik* exodus would result in more virus spreaders traveling. The combination of the Delta variant with human mobility and an uncontrollable situation results in an environment suitable for the spread and mutation of viruses [14].

One of the favorite provinces as a mudik destination is West Java [15], a neighboring province situated southeast of Jakarta. As the home of 48.6 million inhabitants, West Java comprises 27 cities/districts, having had more than 30 thousand active cases by April 2021 [16], far below the cases recorded in the capital city of Jakarta. Every year, the exodus can move more than 20% of the total population of Jakarta into West Java, entering various cities and districts [17]. The mass movement of people from Red Zone Jakarta to West Java is believed to affect the spread of COVID-19 recorded in regions of West Java.

Available Datasets

The Public Health Office of West Java provided the data for COVID-19 spread for each city/district, which comprises three time-series datasets: the number of active cases, recovered, and individuals deceased due to COVID-19 infections [16]. The data interval ranged from early March 2020 until October 2021. In this case study, for May 2021, only the data until the time of the mass exodus for Eid al-Fitr on May 13th, 2021 were considered as representing the spread. The data used to estimate the parameters in model (1) will be explained further in the next section.

Further, the survey of the *mudik* preferences of Indonesian citizens was provided. In brief, the survey collected the preferences of the citizens as to whether they were going to conduct *mudik* or not. If yes, the origin and destination regions were recorded [18]. The number of participants in this survey was ∼62,000, and was distributed mostly to Java citizens. The survey data were intended to estimate the origin–destination matrix for the Eid al-Fitr 2021 exodus. The origin–destination matrix is a two-dimensional matrix that contains information about the magnitude of the movement between locations (zones) within a certain area [19].

MATHEMATICAL MODELING

Generic SIQRD Model

The compartment-approach model is based on the generic model first introduced by Kermack [20], but was modified by accommodating vaccination, quarantine, and reinfection aspects. Thousands of papers related to the model involving several

aspects on the spread of COVID-19 have been published. Several papers using data from Indonesia, i.e., discussing a simple model using data on the city of Jakarta at the beginning of the pandemic using the Richard curve [21, 22], the SIR model [23], the super-spreading problem with Batam–Depok data [24], PSBB in Jakarta [25], scenarios of vaccine implementation in Indonesia [26], and how the model is applied to policymakers [27]. The term SIQRD stands for susceptible, infected, quarantined, recovered, and deceased, classifying people into five groups. The detailed information for each compartment is given in **Table 1**.

Susceptible individuals might have become infected after having had contact with infected individuals, which resulted in a portion of them being moved to compartment I. The addition of compartment E that represents exposed, or carrier, individuals is common to acquire the latency period of diseases [28, 29]. Another approach is to simply use the delay term to accommodate the incubation delay period [30]. However, this study focuses on the quarantine effect and hence drops the latency effect but adds the quarantine compartment. The quarantine rate governs how many of the infected individuals are to be quarantined (hospitalization or self-isolation) and hence, lessens the chance of infecting others. Also, vaccine inoculation is simply accommodated by allowing people from the S compartment to move directly to the R compartment after developing viral immunity. This modification reduces the number of susceptible individuals, depending on the vaccine rate and efficacy, which conforms to the primary goal of the vaccination program [31]. In addition, considering the temporary immunity developed after being inoculated or infected by the disease [32], the modification also covers the cases in which recovered/vaccinated individuals are reinfected by the virus, allowing people in R to move back to S, which can be seen in **Figure 1**.

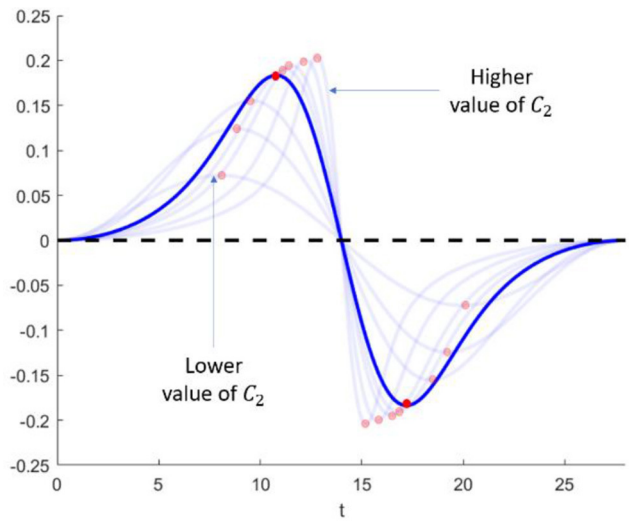


FIGURE 3 | The effect of C_2 change on the location of both peaks.

The mathematical system of the *SIQRD* model is given by:

$$\begin{aligned}
\frac{dS(t)}{dt} &= \pi \hat{N}(t) + \zeta R(t) - \left(\beta \frac{I(t)}{\hat{N}(t)} + \eta v + \mu\right) S(t) \\
\frac{dI(t)}{dt} &= \beta \frac{S(t)\, I(t)}{\hat{N}(t)} - (q + \mu)\, I(t) \\
\frac{dQ(t)}{dt} &= qI(t) - (\gamma + \delta + \mu)\, Q(t) \\
\frac{dR(t)}{dt} &= \gamma Q(t) + \eta v S(t) - (\zeta + \mu)\, R(t) \\
\frac{dD(t)}{dt} &= \delta Q(t)
\end{aligned} \tag{1}$$

where $\hat{N}(t) = N - D(t)$, which represents the number of living individuals over time. The population size is assumed to

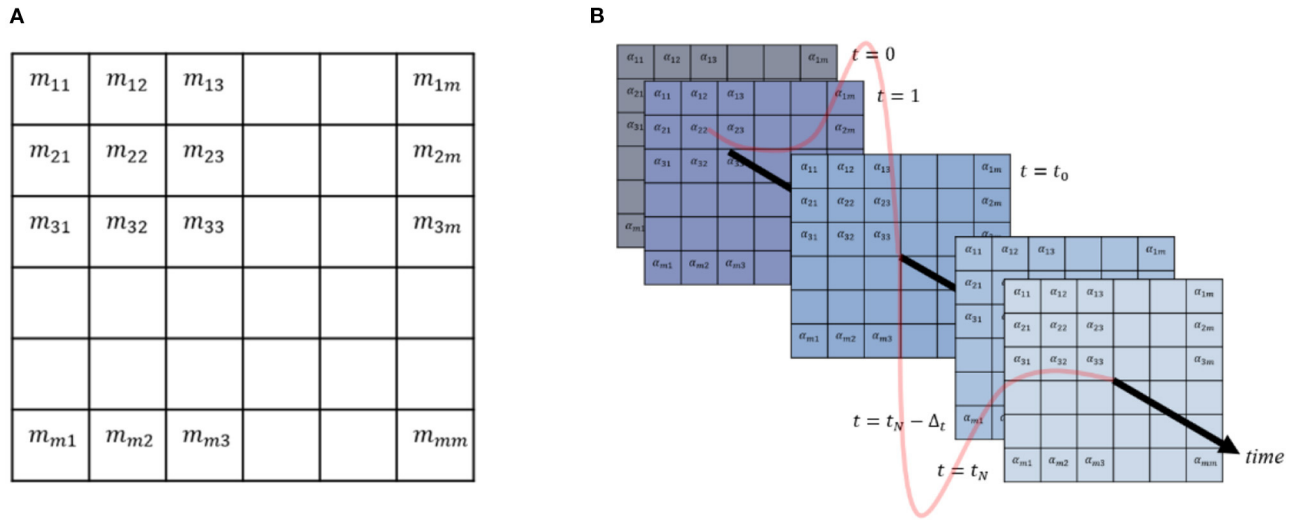


FIGURE 4 | Applying **Algorithm 1** to transform the OD matrix data into the time-dependent mobility rates: **(A)** Origin-Destination matrix; **(B)** time-dependent mobility rates.

remain constant over time and, hence, $\frac{dN}{dt} = 0$ by adjusting $\pi = \mu$. The detailed study of the model is provided in **Supplementary Materials**.

Multiregional SIQRD Model With Mobility Aspect

The multiregional SIQRD, denoted by $\{SIQRD\}_k$, is a generalization of the generic model by simply adding subscript k to each compartment, meaning that the observation of every state dynamic in m regions simultaneously is allowed. The concept of model stratification is commonly used in the analysis of disease spread in relation to region-related observation.

To capture the effect of mass human mobility on the spread of the virus, the multiregional model accommodates the transfer of individuals from one region to another during the exodus. Several strict assumptions should hold to reduce the complexity: only susceptible, inapparently infected, and immune individuals can cross borders. Infected individuals who develop symptoms are banned amid border checks. In contrast, inapparently infected individuals are assumed to travel and pass the border checks easily. The assumption of human mobility considered in this model is that holy days can drive millions of people to travel from one region to another. However, daily mobility is neglected due to its insignificant figure relative to that on holy days. This exodus usually extends no longer than a month, and the travelers are expected to return to their original locations. The infection record is registered in the current place, regardless of the patient's domicile.

It is undeniable that the travel process, including the type of accommodation, plays a significant role in increasing the chance of infections. However, adding this aspect to the model would have increased the complexity and, hence, was not considered. Nevertheless, allowing the human transfer from one region to another is still prospective on capturing the effect of the exodus without taking its travel process into account.

Mathematically, the multiregional SIQRD model, which accommodates interregional mobility, is given by system (2). The multiregional model generalizes the generic one but adds index k and human mobility features.

$$\begin{aligned}
\frac{dS_k(t)}{dt} &= \pi \hat{N}_k(t) + \zeta R_k(t) - \left(\beta_k \frac{I_k(t)}{\hat{N}_k(t)} + \eta \nu_k(t) + \mu\right) S_k(t) \\
&\quad + \sum_{j\neq k}^{m} \left(\alpha_{jk} S_j(t)\right) - \sum_{j\neq k}^{m} \left(\alpha_{kj}\right) S_k(t) \\
\frac{dI_k(t)}{dt} &= \beta_k \frac{I_k(t) S_k(t)}{\hat{N}_k(t)} - \left(q+\mu\right) I_k(t) + \sum_{j\neq k}^{m} \left(\alpha_{jk} I_j(t)\right) \\
&\quad - \sum_{j\neq k}^{m} \left(\alpha_{kj}\right) I_k(t) \qquad (2) \\
\frac{dQ_k(t)}{dt} &= qI_k(t) - (\gamma_k + \delta_k + \mu) Q_k(t) \\
\frac{dR_k(t)}{dt} &= \gamma_k Q_k(t) + \eta \nu_k(t) S_k(t) - (\zeta + \mu) R_k(t) \\
&\quad + \sum_{j\neq k}^{m} \left(\alpha_{jk} R_j(t)\right) - \sum_{j\neq k}^{m} \left(\alpha_{kj}\right) R_k(t) \\
\frac{dD_k(t)}{dt} &= \delta_k Q_k(t)
\end{aligned}$$

TABLE 2 | Estimated rate of transmission, recovery, and deaths due to COVID-19 in Jabodetabek and Jabaraya (district/city scale).

District/city	β	γ	δ	I_0
Jakarta	0.4106	0.1400	0.0021	2,020
Bekasi, the district	0.4087	0.0474	0.0001	105
Bogor, the district	0.4077	0.0387	0.0001	59
Tangerang, the district	0.4027	0.1240	0.0028	80
Bekasi the city	0.4082	0.0716	0.0001	189
Bogor, the city	0.4113	0.0525	0.0001	45
Depok, the city	0.4116	0.0504	0.0001	149
South Tangerang	0.4093	0.1320	0.0032	35
Tangerang, the city	0.4042	0.1213	0.0018	58
West Bandung District	0.4253	0.0900	0.0011	31
Bandung, the district	0.3991	0.0935	0.0001	55
Sumedang District	0.4421	0.0485	0.0001	7
Bandung, the city	0.3814	0.0412	0.0001	359
Cimahi City	0.4341	0.0140	0.0037	11

The population size of each region is not conserved due to the mass mobility from one region to another. However, the sum of all populations in all observed regions remains constant by assuming that no other regions are involved. The initial conditions for variables Q, R, and D are known from the data, yet those for I are estimated.

According to **Table 1**, it is clear that some parameters were assumed to depend on regions, meaning their values represented specific occurrences in each region. The transmission, recovery, and death rates of COVID-19 depended on the provided data and, hence, are written with a subscript. The least-square method was used to estimate these parameters by the input of the provided data. Practically, β, γ, δ, and I_0 were chosen so that the rendered model would be close to the actual data. However, other parameters were considered to be more general and, hence, were assumed to be uniform for each region.

However, the mobility rate α_{jk} indicates the proportion of people traveling from one region to another during Eid al-Fitr. This term was not only assumed to be a region-specific parameter but was also assumed to vary each time, meaning mass mobility would occur occasionally and, hence, have its dynamic. To be depicted by system (2), the term α_{jk} affects only the three variables: S_k, I_k, and R_k. The additional term $\sum_{j\neq k}^{n} (\alpha_{jk} S_j)$ represents the number of susceptible individuals who enter region k from the rest of the regions. However, the term $\sum_{j\neq k}^{n} \left(\alpha_{kj}\right) S_k$ depicts the number of susceptible individuals who leave region k to other regions. Therefore, subtracting the two results in the net change in the susceptible number in the region due to mass exodus. It is noted that the population size in each region may not be conserved, but the sum in all observed regions remains constant each time.

Method of Estimating the Mobility Rate

According to the constructed model, the mobility aspect is accommodated by parameter α_{jk}, representing the daily

TABLE 3 | Estimated OD matrix in Jabodetabek (district/city scale).

		Origin								
		DKI	**Bekasi district**	**Bogor district**	**Tangerang district**	**Bekasi city**	**Bogor city**	**Depok city**	**South Tangerang city**	**Tangerang city**
Destination	DKI	-	0.010	0.070	0.040	0.020	0.010	0.005	0.015	0.005
	Bekasi district	0.015	-	0.010	0.005	0.005	0.040	0.010	0.010	0.010
	Bogor district	0.005	0.015	-	0.005	0.005	0.005	0.040	0.010	0.005
	Tangerang district	0.015	0.010	0.005	-	0.010	0.005	0.005	0.005	0.010
	Bekasi city	0.005	0.005	0.005	0.005	-	0.005	0.005	0.010	0.005
	Bogor city	0.020	0.020	0.010	0.005	0.005	-	0.010	0.005	0.005
	Depok city	0.070	0.020	0.015	0.005	0.005	0.005	-	0.005	0.010
	South Tangerang city	0.050	0.020	0.010	0.005	0.005	0.030	0.005	-	0.020
	Tangerang city	0.010	0.020	0.010	0.010	0.010	0.010	0.005	0.010	-

TABLE 4 | Estimated OD matrix in Jabaraya (district/city scale).

		Origin					
		DKI	**KBB**	**Bandung district**	**Cimahi district**	**Bandung city**	**Cimahi city**
Destination	DKI	-	0	0.022	0.001	0.052	0
	KBB	0	-	0	0	0.001	0
	Bandung district	0.042	0.021	-	0.01	0.101	0.012
	Sumedang district	0.003	0.008	0.005	-	0.012	0.005
	Bandung city	0.027	0.014	0.032	0.31	-	0.01
	Cimahi city	0	0.021	0	0.001	0.005	-

proportion of people traveling between region j and region k. Notice that this parameter is time-dependent and can be either positive or negative. Practically, when α_{jk} is positive, it represents the condition under which a proportion of people travel from region j to region k. The greater this value is, the greater the number of people traveling during the *mudik* occasion. On the other hand, a negative value of α_{jk} represents travel back from region k to region j.

In addition to the quantity aspect, the general behavior of α_{jk} is adjusted to align with the real phenomena: increasing toward Eid al-Fitr and peaking right before the holy day, followed by the people's return in a similar pattern but in reverse. The dynamics are expected to have a significant value near the holy day but almost zero at other times. By the given assumption, the Dawson integral function represents the daily proportion of the population in a specific region traveling during the exodus.

Based on [33], the general form of the Dawson integral function is given as follows:

$$D(t) = -e^{-t^2} \int_0^t e^{x^2} dx \tag{3}$$

Figure 2A depicts the basic graph of the Dawson integral function equipped with the preferred general behavior: positively increasing and reaching its peak toward the holy day, which represents the mass exodus, and then, followed with the reverse pattern in negative values, the return. Moreover, during the holy day, it is assumed that there is no mobility recorded. However, this function has fixed peaks and amplitude as well as thick tails. In order to align with the expected behavior of the mobility rate, the modified Dawson integral function is proposed by adding four additional parameters: C_1, C_2, t_0, and t_1.

$$\alpha_{jk} = \begin{cases} -C_1 \cdot (t) \cdot D\left(C_2\left(t - t_0\right)\right), & \text{for } t \in [t_0 - t_1, t_0 + t_1] \\ 0, & \text{for } t \text{ otherwise} \end{cases} \tag{4}$$

where $(t) = (t - t_0 + t_1)^2 (t - t_0 - t_1)^2$ represents a magnitude adjuster and $D(t)$ is the Dawson integral function. With the given modification, several features of the dynamics can be altered:

1. Adjustment of the amplitude: the generic function is multiplied with C_1 so that the amplitude will be as high as C_1;
2. Adjustment of the peaks' locations, which is accommodated by adding C_2. On the generic Dawson function, the locations of the peaks are fixed, and thus adding the parameter C_2 will adjust the positions of both peaks of the mobility rate;
3. Adjustment of the date of the holy day is accommodated by adding the parameter t_0. According to the modified Dawson function, it is clear that when $t = t_0$, the rate of mobility will be zero. This is relevant due to the assumption that there is no mobility on the day of Eid al-Fitr;
4. Adding the parameter t_1 affects the initial time of the mobility. This adjustment is based on the fact that the generic Dawson function has a thick tail as the time reaches $-\infty$ and ∞. Thus, adding the parameter t_1 to the model can adjust the time at which the mobility is first identified.

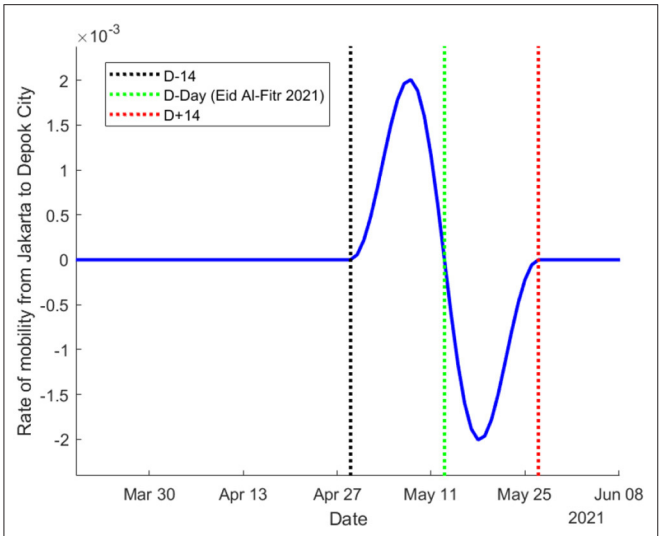


FIGURE 5 | The estimated dynamics of mobility rate from Jakarta to Depok City, ranging from mid of March until late June 2021.

Figure 2 compares the general behavior of the generic Dawson function with the modified one. Note that the modification yields the flexibility of several features. As depicted in **Figure 2B**, both peaks can be adjusted for location and amplitude, allowing for easier adjustment to follow the region-specific phenomena. Otherwise, both tails are already thin, meaning that mobility is considered to happen only near the holy day.

The complexity is now in estimating all of the additional parameters. Parameter t_0 indicates the occurrence of the holy day, which is determined based on the initial data provided. Since the data were available from March 27th, 2020, then t_0 is assumed to be $t_0 = 407$, indicating that Eid al-Fitr 2021 will be held 407 days from the first simulation. Otherwise, t_1 represents the first day near the holiday when people start to travel. Assuming that people start to travel for Eid al-Fitr 14 days before the holy day, $t_1 = 14$.

On determining the value of C_2, **Figure 3** gives insights on how the function changes when the value of C_2 is altered. By adjusting the rest of the parameters to be fixed, the change in C_2 significantly affects the location of both peaks. Increasing the value of C_2 does shift the location of the peaks toward the holy day, and vice versa. This fact indicates that the greater the value, the closer the peak of the mobility is to the holy day. According to [34], it is commonplace for most Indonesians to start to travel 2 days before the holy day, resulting in the mobility peak near this time and packing the roads with extended traffic jams. With the given information, the value of C_2 is determined by forcing the peaks to occur 2 days prior to and after the holy day.

Lastly, parameter C_1 is obtained from the provided dataset since it governs how many people participate in the travel. The data used in the estimation comprise the origin–destination (OD) matrix, which was obtained from the phone survey [18]. Note that the elements in the OD matrix represent the proportion of the population traveling due to the holy day. This quantity

Algorithm 1 : Estimating C_1 given the information from the OD matrix and three additional parameters.

Input Elements of the OD matrix (m_{jk}) and the additional parameters (t_0, t_1, and C_2)
Output C_1: Amplitude adjuster
Steps
1 For arbitrary initial guess C_1, formulate the α_{jk} by means of the modified Dawson integral function ();
2 Evaluate $\int_{t_0-t_1}^{t_1} \alpha_{jk}(t)\,dt$ as a representation of the proportion of the population participating in the exodus during interval $[t_0 - t_1, t_0]$;
3 Solve the optimization problem.

$$\underset{C_1 \in D}{\text{Min}} \left(\left(m_{jk} - \int_{t_0-t_1}^{t_0} \alpha_{jk}(t)\,dt \right)^2 \right)$$

is a single value and is not associated with the daily mobility. The mobility rate α_{jk}, however, depicts the same thing but is time dependent. Henceforth, integrating α_{jk} of time in the range of $[t_0 - t_1, t_0]$ yields exactly the same representation as that of the elements of the OD matrix, i.e., the proportion of a specific region's population traveling from region j to region k. When provided with the information for the other three parameters, C_1 can be obtained by forcing the $\int_{t_0-t_1}^{t_1} \alpha_{jk}(t)\,dt$ to be close to the m_{jk} data. The detailed information for estimating C_1 is given in **Algorithm 1**.

Intuitively, **Algorithm 1** decomposes the single-value data into time-dependent dynamics of mobility rate but conserves their total. For instance, the information from m_{12} is decomposed into $\alpha_{12}(t)$, which has the feature of the modified Dawson integral function but the total $\int_{t_0-t_1}^{t_0} \alpha_{12}(t)\,dt$ still conforms to m_{12}. **Figure 4** shows the visual interpretation of decomposing the m_{jk} data into time-dependent $\alpha_{jk}(t)$.

RESULTS

This section provides information related to the estimation of all parameters and the constructed model's simulation on understanding the effect of the mass mobility on the recorded active cases. According to **Table 1**, there are parameters that are assumed to be constant and independent of regions. Those mentioned parameters are reinfection rate, quarantine rate, vaccine efficacy, and the natural death and birth rates. Those parameters should be first identified by utilizing the existing findings to run the simulations.

According to Goldberg et al. [32], individuals who have recovered after being infected can be reinfected by the virus after 1–2 years, even though no further research has been concluded. The quarantine rate is qualified as the unobservable parameters and, hence, will be assumed to follow the assumption introduced in [35]. The efficacy of vaccines depends on the vaccine brand being delivered. Since Sinovac is being administered to most Indonesians, the vaccine efficacy is assumed to be as high as 57% [36].

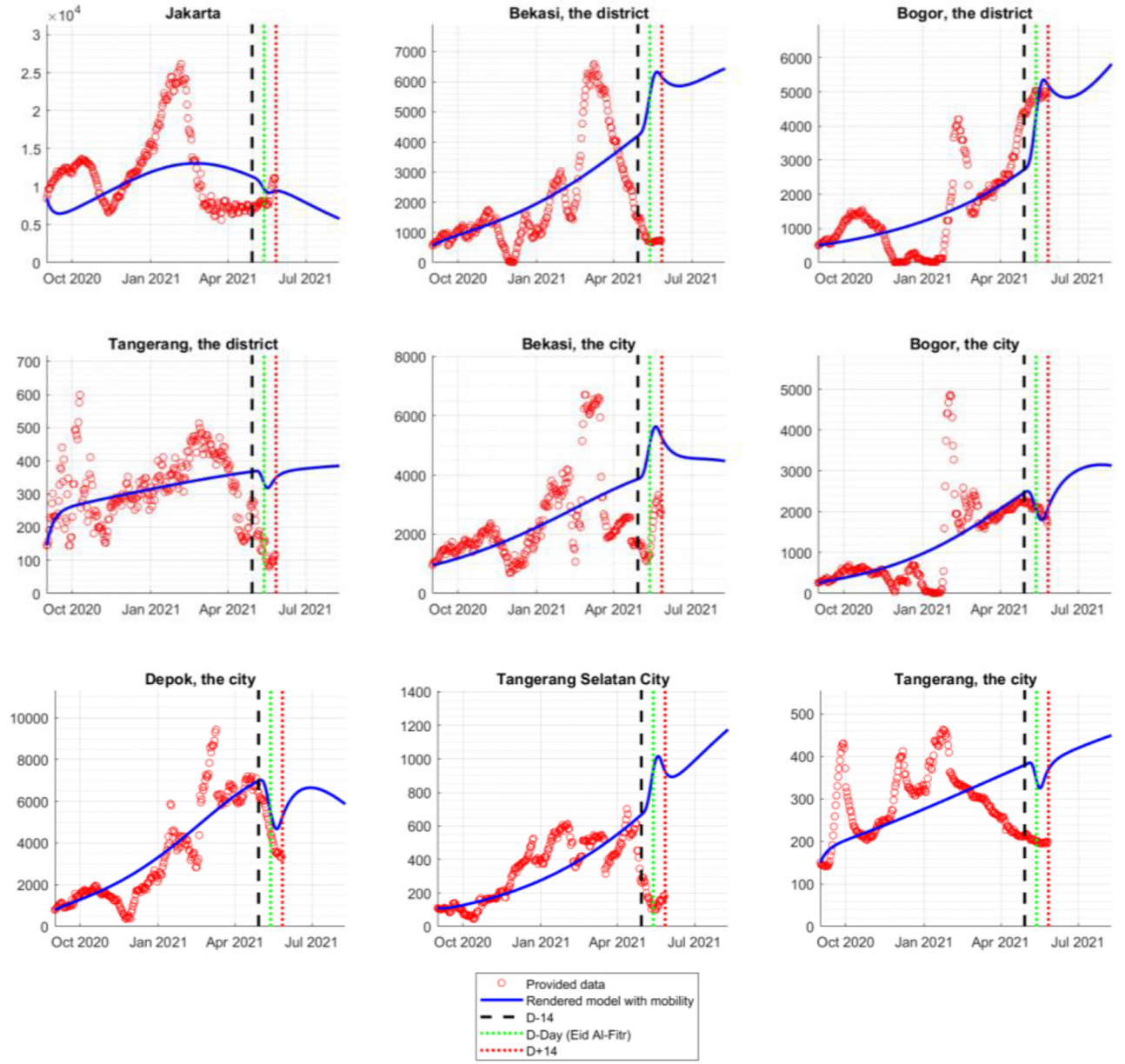


FIGURE 6 | Comparison of the actual data and the rendered model in Jabodetabek during the mass exodus of 2021.

The natural death and birth rates are equal and follow the values introduced in [35] to maintain the constant population size.

Parameter Estimation

Given the data on active cases, cumulative recovery, and deaths for all observed regions of Jabodetabek (Jakarta, Bogor, Depok, Tangerang, and Bekasi) and Jabaraya (Jakarta and Bandung Raya), and assuming that the rate of transmission, recovery, and death due to COVID-19 are constant, the single-value estimated parameters are given in **Table 2**.

Estimated Mobility Rate

According to the phone survey conducted nationally by [18], the results were dominated by respondents from West Java (33.80%), Jakarta (30.95%), Banten (13.07%), Central Java (8.26%), East Java (6.41%), and Yogyakarta (1.61%), and the rest were from other provinces in Indonesia. The survey provided information on how many people in the sample intended to travel during Eid al-Fitr. Assuming that the sample represented the population, the estimation of the OD matrix could be obtained, as shown in **Tables 3**, **4**.

The values originating from Jakarta to West Java were the highest, which means that most of the mobility depicted from the

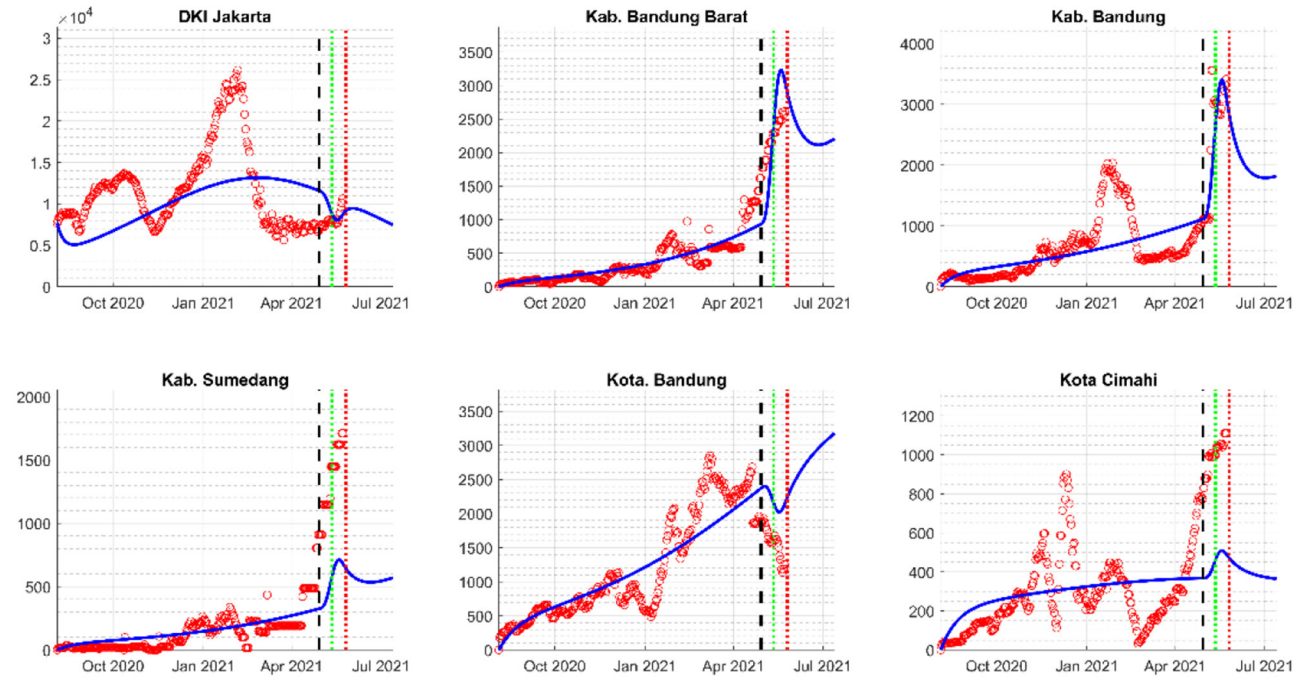


FIGURE 7 | Comparison of the actual data and the rendered model in Jabaraya during the mass exodus of 2021.

survey was from Jakarta to West Java. However, some elements were estimated as zeros. This is a direct implication of the values coming from the phone survey, which was limited to the recorded sample.

By implementing **Algorithm 1**, the time-dependent mobility rate having modified-Dawson-like behavior can be obtained. For visualization, this parameter is shown in **Figure 5**, representing the mobility rate from Jakarta to Depok City. As expected, the effect of mobility started 14 days before the holy day and vanished 14 days later. In the time interval between the 14 days prior to the holy day and the holy day itself, the mobility rate was positive, depicting the travel from Jakarta to Depok City. However, after the holy day, the rate of transmission was negative, representing the return of those who traveled.

Project Active Cases With Mobility Aspect

In the previous section, the constant parameters representing the rate of transmission, recovery, and deaths were obtained from the existing data. The obtained parameters were utilized to evaluate the short-term and long-term projected numbers of active cases of COVID-19. The short-term projection was demonstrated in two scenarios: no exodus or with exodus. Henceforth, the direct effect of mass mobility on the dynamic of active cases can be analyzed.

In Jabodetabek, **Figure 6** demonstrates the projected number of active cases in nine cities/districts. That being said, the estimated β, γ, and δ are assumed to be constant, and hence, the model is not able to capture the fluctuating historical data. However, the analysis mainly focused on the projections that occurred 14 days before and after the Eid.

In the observed interval, the consideration of the mass mobility directly affected the dynamics of the active cases, which can be seen in the interval bracketed by the black and red dashed lines. The four regions, namely Bekasi District, Bogor District, Bekasi City, and Tangerang Selatan City, experienced a significant surge of COVID-19 cases during the mass exodus, while the rest experienced a decline. Specifically, in Depok City, Bogor District, and Bogor City, the rendered model represented the actual data during the mass exodus. In other regions, even though the model seemed to overestimate the actual data, the general trend of the actual data could be depicted in the model, except for Bekasi District, Tangerang District, and Tangerang City.

Figure 7 depicts the simulations given for the regions of Jabaraya. Overall, the rendered model could predict how the active cases would spike or decline once the OD matrix data were provided. The results for Sumedang District and Cimahi City do underestimate the actual data, but the general behavior is well-captured by the data.

Additionally, most of the big or agglomerated cities in the observed regions experienced a decline in the number of active cases, e.g., Jakarta, Bogor City, Tangerang City, and Bandung City. This result confirms that most travelers journeyed from the big city to the more rural areas. On the other hand, other regions experienced a spike representing the significant influx of susceptible individuals to those regions. Regions with higher spikes tended to be the rural areas, which provide recreation, e.g., Bandung District and Cimahi City, meaning that people not only visited families and relatives during the holiday but also

visited some recreational places [37]. This is a reasonable reason for government to bring a travel ban aimed at stemming the Mudik [38].

CONCLUSION

This study produced a method of estimating the movement of people by representing mobility in the form of a function. The calculation provided a reasonably good result compared with the actual data. We used a deterministic model; therefore, in general, the simulation result was the average of the possible occurrences of the given data. The results of this study are considered in formulating policies related to mobility restrictions, especially in urban areas that still have the potential to produce a higher number of cases if the spread cannot be controlled. While currently the number of COVID-19 cases in Indonesia is relatively low, the potential for the spread of this disease is still very high, especially with the discovery of new variants. A relaxation that has been massive enough to control the spread and public awareness to enforce health protocols have also begun to decrease. Therefore, one of the efforts that has contributed to helping prevent the spread of COVID-19 is the policy of regulating community mobility, especially in the Java Island area and around the capital city in particular.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://docs.google.com/spreadsheets/d/1Lhv1tsion5IUHzyOXs93QrqiIuYYkHKimFOgVRfpChI/edit#gid=682277249.

AUTHOR CONTRIBUTIONS

NN and KKS: conceptualization and methodology. KKS and WA: software and writing—original draft preparation. NN, DH, and WA: validation, investigation, and supervision. KKS: formal analysis and visualization. NN, DH, and KKS: resources. NN and DH: data curation. NN, KKS, DH, and WA: writing—review and editing. NN: project administration and funding acquisition. All authors contributed to the article and approved the submitted version.

FUNDING

This research was partly funded by Ministry of Transportation R&D 2021 and PMDSU with research grant: 120Q/IT1.C02/TA.00/2021.

ACKNOWLEDGMENTS

The authors thank Nunuj Nurdjanah, S.Si., M.T., from Litbang Jalan dan Perkeretaapian Indonesia.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fams.2022.912150/full#supplementary-material

REFERENCES

1. Oraby T, Tyshenko M, Maldonado J, Vatcheva K, Elsaadany S, Alali W, et al. Modeling the effect of lockdown timing as a COVID-19 control measure in countries with differing social contacts. *Sci Rep*. (2021) 3354:886. doi: 10.1101/2020.11.14.20231886
2. Han E, Lessons learnt from easing COVID-19 restrictions: an analysis of countries and regions in Asia Pacific and Europe. *Lancet*. (2020) 396:1525–34. doi: 10.1016/S0140-6736(20)32007-9
3. CfDCa. *Prevention, Quarantine and Isolation*. New York, NY: CDC (2020).
4. Adibrata R, Iskandar N. *Anies Extends Jakarta PSBB, June Becomes Transition Period, beritajakarta*. BeritaJakarta, Jakarta (2020).
5. Afifa L. *Jakarta Records Over 3,000 New COVID-19 Cases for Four Consecutive Days*. New York, NY: Tempo Co. (2021).
6. Onishi N. *Indonesians Go Home, by the Millions*. New York, NY: The New York Times (2010).
7. Muhammad HH. *Researchers: Travelers in Mudik 2021. Will Significantly Increase (in Bahasa Indonesia)*. Manila, Philippines: Republika (2021).
8. Mun X, Yeh AG-O, Xhang X. *The Interplay of Spatial Spread of COVID-19 and Human Mobility in the Urban System of China During the Chinese New Year*. Environment and Planning B: Urban Analytics and City Science, Thousand Oaks, CA: SAGE Journals (2020).
9. Syofyan D. *Lebaran and Local Pride in the Annual Mudik Custom*. Jakarta, Indonesia: The Jakarta Post (2105).
10. Yinglun S. *Millions in Bangladesh on the Move as Eid Exodus Begins Amid Beefed up Security*. Beijing, China: XINHUANET (2021).
11. Anggraini AD. *Number of Eid Exodus in Indonesia Continues to Increase (in Bahasa Indonesia)*. Indonesia: Indonesiabaik (2018).
12. Tulus TA. Efektivitas Aplikasi Garut Intan Pada Pemberian Informasi Arus Mudik Dan Arus Balik Di Kawasan Hukum Polres Garut (in Bahasa Indonesia). *Indon J Police Stud*. (2020) 4. Available online at: https://journal.akademikepolisian.com/index.php/ijps/article/view/511
13. Andaka D. Dampak Pelarangan Mudik Akibat Pandemik COVID-19 terhadap Bisnis Angkutan Udara di Indonesia (in Bahasa Indonesia). *J Civil Eng Plann*. (2021) 1:116–29.
14. Oka T. The effect of human mobility restrictions on the COVID-19 transmission network in China, *PLoS ONE*. (2021) 16:e0254403. doi: 10.1371/journal.pone.0254403
15. Veruswati M. Correlation between Local Eid-al-Fitr Homecoming (Mudik) with coronavirus disease-19 during ramadhan season amidst large-scale social distancing in Indonesia. *Maced J Med Sci*. (2021) 8:570–3. doi: 10.3889/oamjms.2020.5369
16. kawal COVID. *Kawal Informasi Seputar COVID-19 secara tepat dan akurat. (in Bahasa Indonesia)*. Indonesia: KawalCOVID19.id (2021).
17. Firman T. Demographic and spatial patterns of Indonesia's recent urbanisation. *Popul Space Place*. (2004) 10:42–34. doi: 10.1002/psp.339
18. Kemenhub. *Laporan Antara: Penelitian Penyelenggara Angkutan Lebaran tahun 2021*. Jakarta, Indonesia: Ministry of Transportation (2021).
19. Basri H, Jatmika B, Arifi Z, Sundari S, Barkah B. Measurement of accessibility index using gravity model based on GIS in Sukabumi District, in 2018. In: *International Conference on Computing, Engineering, Design (ICCED)* (Bangkok, Thailand) (2018).
20. Kermack W. *A Contribution to the Mathematical Theory of Epidemics*. London, UK: The Royal Society Publishing (1927).
21. Nuraini NK, Sukandar AM. Modeling simulation of COVID-19 in Indonesia based on early endemic data. *Commun Biomath Scie*. (2020) 3:1–8. doi: 10.5614/cbms.2020.3.1.1

22. Darti I, Suryanto A, Panigoro HS. Susanto H. Forecasting COVID-19 epidemic in Spain and Italy using A generalized richards model with quantified uncertainty. *Commun Biomath Sci.* (2020) 3:90–100. doi: 10.5614/cbms.2020.3.2.1
23. Soewono E. On the analysis of covid-19 transmission in Wuhan, diamond princess and Jakarta-cluster. *Commun Biomath Sci.* (2020) 3:9–18. doi: 10.5614/cbms.2020.3.1.2
24. Hasan A, Susanto H, Kasim M, Nuraini N, Lestari B, Triany D, et al. Superspreading in early transmissions of COVID-19 in Indonesia. *Sci Rep*. (2020) 10:22386. doi: 10.1101/2020.06.28.20142133
25. Hasan A, Nasution Y, Susanto H, Putri E, Tjahjono V, Puspita D, et al. Modeling COVID-19 transmissions and evaluation of large-scale social restriction in Jakarta, Indonesia. *medRxiv*. (2020). doi: 10.1101/2020.10.30.20222984
26. Fuady A, Nuraini N, Sukandar KK, Lestari BW. Targeted vaccine allocation could increase the COVID-19 vaccine benefits amidst its lack of availability: a mathematical modeling in Indonesia. *Vaccines.* (2021) 9:462. doi: 10.3390/vaccines9050462
27. Hasan A, Putri E, Susanto H, Nuraini N. Data-driven modeling and forecasting of COVID-19 outbreak for public policy making. *ISA Trans.* (2021) 124. doi: 10.1016/j.isatra.2021.01.028
28. Hattaf K, Mohsen AA, Harraq J, Achtaih N. Modeling the dynamics of COVID-19 with carrier effect and environmental contamination. *World Sci.* (2021) 12:483. doi: 10.1142/S1793962321500483
29. Jaouimaa FZ, Dempsey D, Van Osch S, Kinsella S, Burke K, Wyse J, et al. An age-structured SEIR model for COVID-19 incidence in Dublin, Ireland with framework for evaluating health intervention cost. *PLoS ONE.* (2021) 16:e0260632. doi: 10.1371/journal.pone.0260632
30. Liu L. A delayed SIR model with general. *Adv Diff Eq.* (2015) 329:1–11. doi: 10.1186/s13662-015-0619-z
31. Swan DA, Bracis C, Janes H, Moore M, Matrajt L, Reeves DB, et al. COVID-19 vaccines that reduce symptoms but do not block infection need higher coverage and faster rollout to achieve population impact. *Sci Rep.* (2021) 11:15531. doi: 10.1038/s41598-021-94719-y
32. Goldberg Y, Mandel M, Bar-On YM, Bodenheimer O, Freedman L, Haas EJ, et al. Waning immunity after the BNT162b2 Vaccine in Israel. *N Engl J Med*. (2021) 385:e85. doi: 10.1056/NEJMoa2114228
33. Temme N. *Error Fucntions, Dawson's and Fresnel Integrals*. Cambridge: Cambridge University Press (2010).
34. Rosana FCR. *Kemenhub Prediksi Puncak Arus Balik Lebaran pada Sabtu dan Minggu*. New York, NY: TEMPO (2021).
35. Nuraini N, Sukandar KK, Hadisoemarto P, Susanto H, Hasan AI, Sumarti N. Mathematical models for assessing vaccination scenarios in several provinces in Indonesia, *J Infect Dis Model.* (2021) 6:1236–58. doi: 10.1016/j.idm.2021.09.002
36. IfHMa. *Evaluation, COVID-19 Vaccine Efficacy Summary*. Seattle, Washington: IHME (2021).
37. Perdana R. *Warga Kota Bandung Boleh Mudik ke Kabupaten dan Cimahi, tapi Tidak ke Jatinangor (in Bahasa Indonesia)*. Bandung, Indonesia: Pikiran Rakyat. (2021).
38. Yuniar RW. Indonesia risks India-style Covid-19 surge as millions skirt Eid ul-Fitr travel ban. *Coronav Pandemic*. (2021). Available online at: https://www.scmp.com/week-asia/health-environment/article/3132706/indonesia-risks-india-style-covid-19-surge-millions

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Applied Mathematics and Statistics

ORIGINAL RESEARCH
published: 16 June 2022
doi: 10.3389/fams.2022.912674

The Application of the Functional Variable Method for Solving the Loaded Non-linear Evaluation Equations

Bazar Babajanov[1] *and Fakhriddin Abdikarimov*[2]*

[1] *Department of Applied Mathematics and Mathematical Physics, Urgench State University, Urgench, Uzbekistan,* [2] *Khorezm Mamun Academy, Khiva, Uzbekistan*

In this article, we construct exact traveling wave solutions of the loaded Korteweg-de Vries, the loaded modified Korteweg-de Vries, and the loaded Gardner equation by the functional variable method. The performance of this method is reliable and effective and gives the exact solitary and periodic wave solutions. All solutions to these equations have been examined and 3D graphics of the obtained solutions have been drawn by using the Matlab program. We get some traveling wave solutions, which are expressed by the hyperbolic functions and trigonometric functions. The graphical representations of some obtained solutions are demonstrated to better understand their physical features, including bell-shaped solitary wave solutions, singular soliton solutions, and solitary wave solutions of kink type. Our results reveal that the method is a very effective and straightforward way of formulating the exact traveling wave solutions of non-linear wave equations arising in mathematical physics and engineering.

Keywords: the loaded Korteweg-de Vries equation, the loaded modified Korteweg-de Vries equation, periodic wave solutions, soliton wave solutions, the loaded Gardner equation, functional variable method

AMS Subject Classification: 34A34, 34B15, 35Q51, 35J60, 35J66.

OPEN ACCESS

Edited by:
Tri Nguyen-Huu,
Institut de Recherche pour le Développement (IRD), France

Reviewed by:
Haci Mehmet Baskonus,
Harran University, Turkey
Omar Abu Arqub,
Al-Balqa Applied University, Jordan

***Correspondence:**
Fakhriddin Abdikarimov
goodluck_0714@mail.ru

Specialty section:
This article was submitted to Mathematics of Computation and Data Science, a section of the journal Frontiers in Applied Mathematics and Statistics

Received: *04 April 2022*
Accepted: *09 May 2022*
Published: *16 June 2022*

Citation:
Babajanov B and Abdikarimov F (2022) The Application of the Functional Variable Method for Solving the Loaded Non-linear Evaluation Equations. Front. Appl. Math. Stat. 8:912674. doi: 10.3389/fams.2022.912674

1. INTRODUCTION

The investigation of exact traveling wave solutions to non-linear evolutions equations plays an important role in the study of non-linear physical phenomena. These equations arise in several fields of science, such as fluid dynamics, physics of plasmas, biological models, non-linear optics, chemical kinetics, quantum mechanics, ecological systems, electricity, ocean, and sea. One of the most important non-linear evolution equations is Korteweg De Vries (KdV) equation.

The KdV equation was first observed by John Scott Russell in experiments, and then Lord Rayleigh and Joseph Boussinesq studied it theoretically. Finally, in 1895, Korteweg and De Vries formulated a model equation to describe the aforementioned water wave, which helped to prove the existence of solitary waves. In the mid-1960s, Zabusky and Kruskal discovered the remarkably stable particle-like behavior of solitary waves. According to their study, solitary waves described by the KdV equation can pass through each other keeping their speed and shape unchanged. As a result, the name "soliton" is defined. In the wake of these discoveries, solitary wave theory boosted the development of many areas of science and technology. After 100 years, integrable systems

developed deeply and soliton theory was widely applied in many areas. The KdV equation

$$u_t + 6uu_x + u_{xxx} = 0, \tag{1}$$

has many connections to several branches of physics. The Equation (1) is especially important due to the potential application of different properties of electrostatic waves in the development of new theories of chemical physics, space environments, plasma physics, fluid dynamics, astrophysics, optical physics, nuclear physics, geophysics, dusty plasma, fluid mechanics, and different other fields of applied physics [1–11].

In recent years, studying electrostatic waves specifically to discuss different properties of solitary waves in the field of soliton dynamics has played a significant role for many researchers and has received considerable attention from them. The ion-acoustic solitary wave is one of the fundamental non-linear wave phenomena appearing in plasma physics. In 1973, Hans Schamel studies a modified Korteweg-de Vries equation for ion-acoustic waves which is expressed in the following basic form

$$u_t + 6u^2u_x + u_{xxx} = 0. \tag{2}$$

This equation has been applied widely, e.g., in the molecular chain model, the generalized elastic solid, and so on [12–14]. Non-linear interactions between low-hybrid waves and plasmas can be described well by using the mKdV equation [15].

In 1968, the Gardner equation is an integrable non-linear partial differential equation introduced by the mathematician Clifford Gardner to generalize the KdV equation and modified the KdV equation. This equation can be written in a normalized form as follows:

$$u_t + 2\alpha uu_x + 3\beta u^2u_x + u_{xxx} = 0. \tag{3}$$

If the coefficient $\beta > 0$, Equation (3) admits two families of solitons and oscillating wave packets (called breathers), whereas if $\beta < 0$, only one category of solitons exists [16]. The Gardner equation plays an important role in various branches of physics, such as plasma physics, fluid physics, and quantum field theory [17, 18]. It also describes a variety of wave phenomena in plasma and solid state physics [19, 20].

In arterial mechanics, a model is widely used in which the artery is considered as a thin-walled prestressed elastic tube with a variable radius (or with stenosis) and blood as an ideal fluid [21]. The governing equation that models weakly non-linear waves in such fluid-filled elastic tubes is the modified Korteweg–de Vries equation

$$u_t - 6u^2u_x + u_{xxx} - h(t)u_x = 0,$$

where t - is a scaled coordinate along the axis of the vessel after static deformation characterizing axisymmetric stenosis on the surface of the arterial wall. x - is a variable that depends on time and coordinates along the axis of the vessel. $h(t)$- is a form of stenosis and $u(x, t)$ characterizes the average axial velocity of the fluid.

We suppose that a form of stenosis $h(t)$ is proportional to $u(0, t)$, and we consider the loaded KdV, the loaded modified KdV and the loaded Gardner equation

$$u_t - 6\alpha uu_x + u_{xxx} + \gamma(t)u(0,t)u_x = 0, \tag{4}$$

$$u_t - 12\alpha u^2u_x + u_{xxx} + \gamma(t)u(0,t)u_x = 0, \tag{5}$$

$$u_t + 2\alpha uu_x + 3\beta u^2u_x + u_{xxx} + \gamma(t)u(0,t)u_x = 0, \tag{6}$$

where $u(x, t)$ is an unknown function, $x \in R$, $t \geq 0$, α, and β are any constants, $\gamma(t)$ is the given real continuous function.

Many powerful and direct methods have been developed to find special solutions of non-linear evolution equations such as, Weierstrass elliptic function method [22], Jacobi elliptic function expansion method [23], tanh-function method [24], inverse scattering transform method [25], Hirota method [26], Backlund transform method [27], exp-function method [28], truncated Painleve expansion method [29], extended tanh-method [30], and the homogeneous balance method [31] are used for searching the exact solutions.

We establish exact traveling wave solutions of the loaded KdV, the loaded modified KdV, and the loaded Gardner equation by the functional variable method. The performance of this method is reliable and effective and gives the exact solitary wave solutions and periodic wave solutions. The traveling wave solutions obtained *via* this method are expressed by hyperbolic functions and trigonometric functions. The graphical representations of some obtained solutions are demonstrated to better understand their physical features, including bell-shaped solitary wave solutions, singular soliton solutions, and solitary wave solutions of kink type. This method presents wider applicability for handling non-linear wave equations.

In the recent years, the study of the stability of traveling waves of periodic and soliton types associated with non-linear dispersive equations has increased significantly. A rich variety of new mathematical problems have emerged, as well as the physical importance related to them. This subject is often studied in relation to the natural symmetries associated with the model and by perturbations of symmetric classes, e.g., the class of periodic functions with the same minimal period as the underlying wave. In the case of shallow-water wave models, a formal stability theory of periodic and soliton traveling waves has started.

It is known that the loaded differential equations contain some traces of an unknown function. In [32–38], the term "loaded equation" was used for the first time, the most general definitions of the loaded differential equation were given, and also detailed classifications of the differential loaded equations, as well as their numerous applications, were presented. A complete description of solutions of the non-linear loaded equations and their applications can be found in articles [39–45].

2. DESCRIPTION OF THE FUNCTIONAL VARIABLE METHOD

Consider non-linear evolution equations with independent variables $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{t}$ is of the form

$$F(u, u_x, u_y, u_t, u_{xx}, u_{tt}, u_{yy}, u_{xy}, u_{xt}, u_{yt}...) = 0, \tag{7}$$

where F is a polynomial in $u = u(x, y, t)$ and its partial derivatives. In [46, 47], Zerarka and others have summarized the functional variable method in the following:

Step 1. We use the wave transformation

$$\xi = px + qy - kt, \tag{8}$$

where p and q are constants, and k is the speed of the traveling wave.

Next, we can introduce the following transformation for a traveling wave solution of Equation (7)

$$u(x, y, t) = u(\xi), \tag{9}$$

and the chain

$$\frac{\partial u}{\partial x} = p\frac{du}{d\xi}, \frac{\partial u}{\partial y} = q\frac{du}{d\xi}, \frac{\partial u}{\partial t} = -k\frac{du}{d\xi}, ... \tag{10}$$

Using Equations (9) and (10), the non-linear partial differential Equation (7) can be transformed into an ordinary differential equation of the form

$$P(u, u', u'', u''', ...) = 0, \tag{11}$$

where P is a polynomial in $u(\xi)$ and its total derivatives, $u' = \frac{du}{d\xi}$.

Step 2. Then we make a transformation in which the unknown function u is considered a functional variable in the form

$$u' = F(u), \tag{12}$$

then, the solution can be found by the relation

$$\int \frac{du}{F(u)} = \xi + \xi_0,$$

here, ξ_0 is a constant of integration which is set equal to zero for convenience. Some successive differentiations of u in terms of F are given as

$$\begin{aligned} u'' &= \tfrac{dF(u)}{du}\tfrac{du}{d\xi} = \tfrac{dF(u)}{du}F(u) = \tfrac{1}{2}\tfrac{d(F^2(u))}{du}, \\ u''' &= \tfrac{1}{2}\tfrac{d^2(F^2(u))}{du^2}\sqrt{F^2(u)}, \\ u'''' &= \tfrac{1}{2}\left[\tfrac{d^3(F^2(u))}{du^3}F^2(u) + \tfrac{d^2(F^2(u))}{du^2}\tfrac{d(F^2(u))}{du}\right]. \end{aligned} \tag{13}$$

Step 3. The ordinary differential Equation (11) can be reduced in terms of u, F, and its derivatives upon using the expressions of Equation (13) into Equation (7) gives

$$H(u, \frac{dF(u)}{du}, \frac{d^2F(u)}{du^2}, \frac{d^3F(u)}{du^3}, ...) = 0. \tag{14}$$

The key idea of this particular form Equation (14) is of special interest because it admits analytical solutions for a large class of non-linear wave type equations. After integration, Equation (14) provides the expression of F and this, together with Equation (12), give appropriate solutions to the original problem.

3. SOLUTIONS OF THE LOADED KDV EQUATION

We will show how to find the exact solution of the loaded KdV by the functional variable method. Using the wave variable

$$u(x, t) = u(\xi),\ \xi = px - kt,$$

that will convert Equation (4) to an ordinary differential equation

$$-ku' - 6\alpha puu' + p^3u''' + \gamma(t)pu(0, t)u' = 0. \tag{15}$$

Integrating once Equation (15) with respect to ξ, and put the constant of integration zero, we have

$$u'' = \frac{1}{p^3}\left(3\alpha pu^2 + \left(k - \gamma(t)pu(0, t)\right)u\right). \tag{16}$$

Following Equation (13), it is easy to deduce from Equation (16) an expression for the function $F(u)$

$$\frac{1}{2}\frac{d\left(F^2(u)\right)}{du} = \frac{1}{p^3}\left(3\alpha pu^2 + \left(k - \gamma(t)pu(0, t)\right)u\right). \tag{17}$$

Integrating Equation (17) and setting the constant of integration to zero yields

$$F^2(u) = \frac{1}{p^3}\left(2\alpha pu^3 + \left(k - \gamma(t)pu(0, t)\right)u^2\right)$$

$$F(u) = u\sqrt{\frac{2\alpha}{p^2}}\sqrt{u - \frac{\gamma(t)pu(0, t) - k}{2\alpha p}}$$

$$F(u) = u\sqrt{\frac{2\alpha}{p^2}}\sqrt{u - \eta(t)}, \tag{18}$$

where $\eta(t) = \frac{\gamma(t)pu(0,t)-k}{2\alpha p}$. From Equation (12) and Equation (18), we deduce that

$$\frac{du}{u\sqrt{u - \eta(t)}} = \sqrt{\frac{2\alpha}{p^2}}d\xi. \tag{19}$$

After integrating Equation (19), with zero constant of integration, we have the following exact solution

$$u(x, t) = \frac{\gamma(t)pu(0, t) - k}{2\alpha p}\frac{1}{\cos^2\sqrt{\frac{\gamma(t)pu(0,t)-k}{4p^3}}\left(px - kt\right)}. \tag{20}$$

It is obvious that the function $u(0, t)$ can be easily found based on expression (20).

We have several types of traveling wave solutions of the loaded KdV equation as follows:

1) When $\sqrt{\frac{\gamma(t)pu(0,t)-k}{4p^3}} > 0$, we have the periodic wave solution

$$u(x, t) = \frac{\gamma(t)pu(0, t) - k}{2\alpha p}\frac{1}{\cos^2\sqrt{\frac{\gamma(t)pu(0,t)-k}{4p^3}}\left(px - kt\right)}.$$

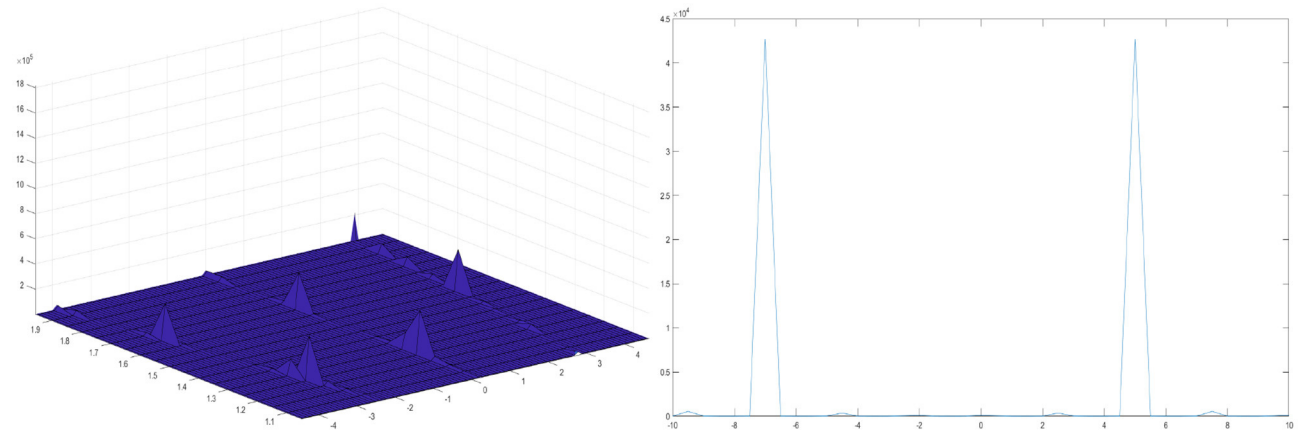


FIGURE 1 | Periodic wave solution of the loaded KdV equation for $k = -1$, $\alpha = 0.5$, $p = 1$, and $\gamma(t) = t$.

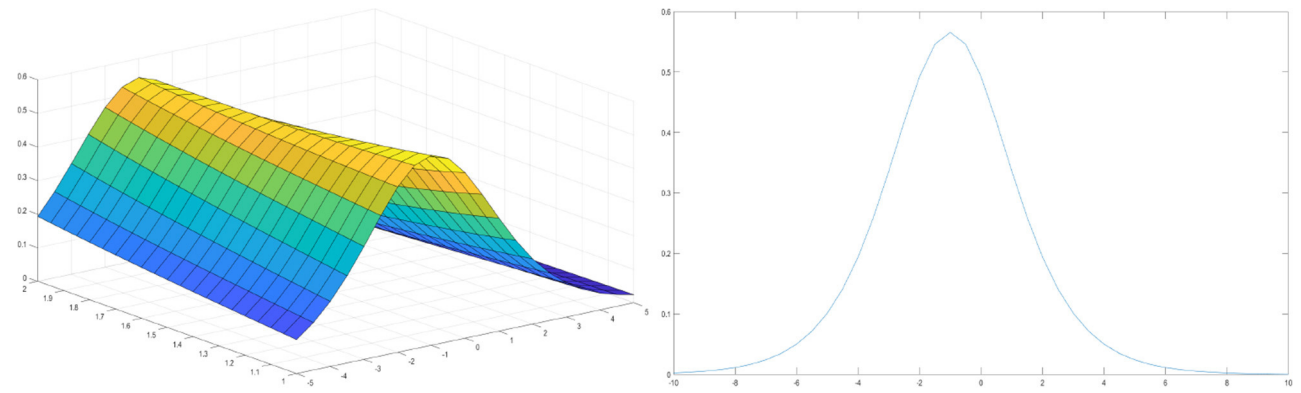


FIGURE 2 | Solitary wave solution of the loaded KdV equation for $k = 1$, $\alpha = 0.5$, $p = -1$, and $\gamma(t) = -t$.

2) When $\sqrt{\frac{\gamma(t)pu(0,t)-k}{4p^3}} < 0$, we have the solitary wave solution

$$u(x,t) = \frac{\gamma(t)pu(0,t)-k}{2\alpha p} \frac{1}{\cosh^2 \sqrt{\frac{\gamma(t)pu(0,t)-k}{4p^3}} (px - kt)}.$$

Now, by choosing free parameters, we will write the traveling wave solutions of the loaded KdV equation in the simple form which can be used for the graphical illustrations.

If $k = -1, \alpha = 0.5, p = 1$ and $\gamma(t) = t$, then we have

$$u(x,t) = \frac{tu(0,t)+1}{\cos^2 \sqrt{\frac{tu(0,t)+1}{4}} (x+t)}, \tag{21}$$

It is obvious that the function $u(0,t)$ can be easily found based on expression (21).

If $k = 1, \alpha = 0.5, p = -1$, and $\gamma(t) = -t$, then we have

$$u(x,t) = -\frac{tu(0,t)+1}{\cosh^2 \sqrt{\frac{tu(0,t)+1}{4}} (t-x)}, \tag{22}$$

It is obvious that the function $u(0,t)$ can be easily found based on expression (22).

4. GRAPHICAL REPRESENTATION OF THE LOADED KDV EQUATION

We have presented some graphs of solitary and periodic waves constructed by taking suitable values of the involved unknown parameters to visualize the underlying mechanism of the original physical phenomena. Using mathematical software Matlab, 3D plots of the obtained solutions have been shown in **Figures 1**, **2**. A soliton or solitary wave in the concept of mathematical physics is defined as a self-reinforcing wave package that retains its shape. It propagates at a constant amplitude and velocity. Solitons are solutions of a common class of non-linearly partially differential equations with weak linearity describing physical systems. The existence of periodic traveling waves usually depends on the parameter values in a mathematical equation. If there is a periodic traveling wave solution, then there is typically a family of such solutions, with different wave speeds. For partial

differential equations, periodic traveling waves typically occur for a continuous range of wave speeds. The physical description of the 3D loaded KdV equation of the installed exact moving wave solutions is discussed in this section. In the physical definition section, 3D surface drawings, contour maps, and 2D drawings of the developed moving wave solutions of the latest 3D loaded KdV equations are discussed. The 3D line plot emphasizes the amount of variability over time or compares multiple wave elements. The wave points were sequentially designed using equal interval breaks and connected by a line to emphasize the relationship of the wave points. Three-dimensional elegance is used to give visual attention to the diagram. Two-dimensional line drawings are used to represent very high and low frequencies and amplitudes.

5. SOLUTIONS OF THE LOADED MODIFIED KDV EQUATION

Assume that Equation (5) has an exact solution in the form of a traveling wave

$$u(x,t) = u(\xi),\ \xi = px - kt,$$

the Equation (5) can be converted to an ordinary differential equation

$$-ku' - 12\alpha p u^2 u' + p^3 u''' + \gamma(t) p u(0,t) u' = 0. \tag{23}$$

Once integrating (23), setting the constant of integrating to zero, we obtain

$$u'' = \frac{1}{p^3}\left(4\alpha p u^3 + \left(k - \gamma(t) p u(0,t)\right) u\right). \tag{24}$$

Following Equation (13), it is easy to deduce from Equation (24) an expression for the function $F(u)$

$$\frac{1}{2}\frac{d\left(F^2(u)\right)}{du} = \frac{1}{p^3}\left(4\alpha p u^3 + \left(k - \gamma(t) p u(0,t)\right) u\right). \tag{25}$$

Integrating Equation (25) with respect to u and after the mathematical manipulations, we have

$$F^2(u) = \frac{1}{p^3}\left(2\alpha p u^4 + \left(k - \gamma(t) p u(0,t)\right) u^2\right)$$

$$F(u) = \frac{u}{p}\sqrt{2\alpha}\sqrt{u^2 - \frac{\gamma(t) p u(0,t) - k}{2\alpha p}}$$

$$F(u) = \frac{u}{p}\sqrt{2\alpha}\sqrt{u^2 - \varphi(t)}, \tag{26}$$

where $\varphi(t) = \frac{\gamma(t) p u(0,t) - k}{2\alpha p}$. From Equation (12) and Equation (26), we deduce that

$$\frac{du}{u\sqrt{u^2 - \varphi(t)}} = \frac{\sqrt{2\alpha}}{p} d\xi. \tag{27}$$

After integrating Equation (27), with zero constant of integration, we have the following exact solution

$$u(x,t) = \sqrt{\frac{\gamma(t) p u(0,t) - k}{2\alpha p}} \frac{1}{\cos\sqrt{\frac{\gamma(t) p u(0,t) - k}{p^3}}\left(px - kt\right)}. \tag{28}$$

It is obvious that the function $u(0,t)$ can be easily found based on expression (28).

We have several types of traveling wave solutions of the loaded modified KdV equation as follows:

1) When $\sqrt{\frac{\gamma(t) p u(0,t) - k}{p^3}} > 0, \alpha > 0$, we have the periodic wave solution

$$u(x,t) = \sqrt{\frac{\gamma(t) p u(0,t) - k}{2\alpha p}} \frac{1}{\cos\sqrt{\frac{\gamma(t) p u(0,t) - k}{p^3}}\left(px - kt\right)}.$$

2) When $\sqrt{\frac{\gamma(t) p u(0,t) - k}{p^3}} < 0, \alpha < 0$, we have the solitary wave solution

$$u(x,t) = \sqrt{\frac{\gamma(t) p u(0,t) - k}{2\alpha p}} \frac{1}{\cosh\sqrt{\frac{\gamma(t) p u(0,t) - k}{p^3}}\left(px - kt\right)}.$$

Now, by choosing free parameters, we will write the traveling wave solutions of the loaded modified KdV equation in the simple form which can be used for the graphical illustrations.

If $k = -1, \alpha = 0.5, p = 1$, and $\gamma(t) = t$, then we have

$$u(x,t) = \frac{\sqrt{tu(0,t) + 1}}{\cos\sqrt{tu(0,t) + 1}\,(x + t)}, \tag{29}$$

It is obvious that the function $u(0,t)$ can be easily found based on expression (29).

If $k = 1, \alpha = -0.5, p = 1$, and $\gamma(t) = -t$, then we have

$$u(x,t) = \frac{\sqrt{tu(0,t) + 1}}{\cosh\sqrt{tu(0,t) - 1}\,(x - t)}, \tag{30}$$

It is obvious that the function $u(0,t)$ can be easily found based on expression (30).

6. PHYSICAL EXPLANATION OF THE LOADED MODIFIED KDV EQUATION

We have shown how to find the solutions of the loaded modified KdV equation in 3D plot formats to make it easier to imagine. Graphical representation is an effective tool for communication and it exemplifies evidently the solutions to the problems. The graphical illustrations of the solutions are depicted in **Figures 3**, **4**. Solitary and periodic wave solutions represent an important type of solutions for non-linear partial differential equations as many non-linear partial differential equations have been found to have a variety of solitary wave solutions. The solitary wave solutions obtained in this article are encouraging, applicable, and could be helpful in analyzing long wave propagation on the surface of a fluid layer under the action of gravity, iron sound waves in plasma, and vibrations in a non-linear string.

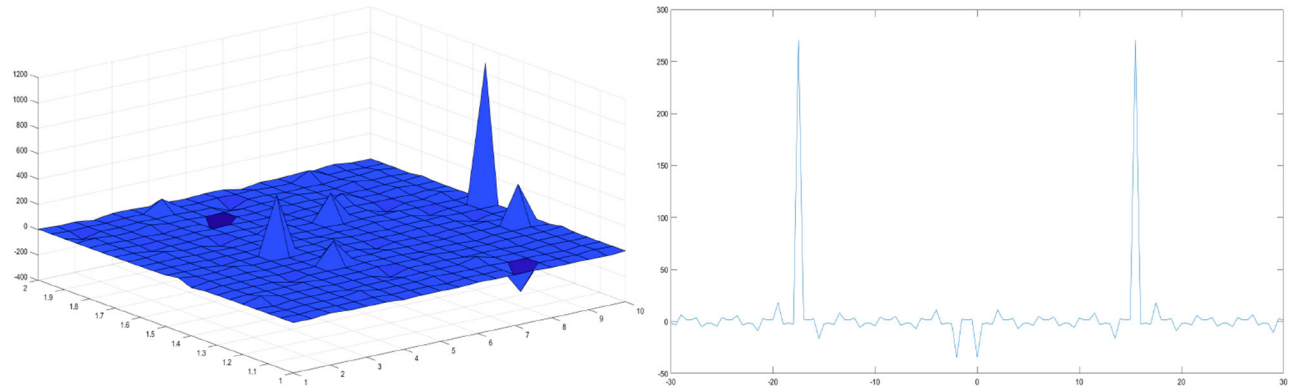


FIGURE 3 | Periodic wave solution of the loaded modified KdV equation for $k = -1$, $\alpha = 0.5$, $p = 1$, and $\gamma(t) = t$.

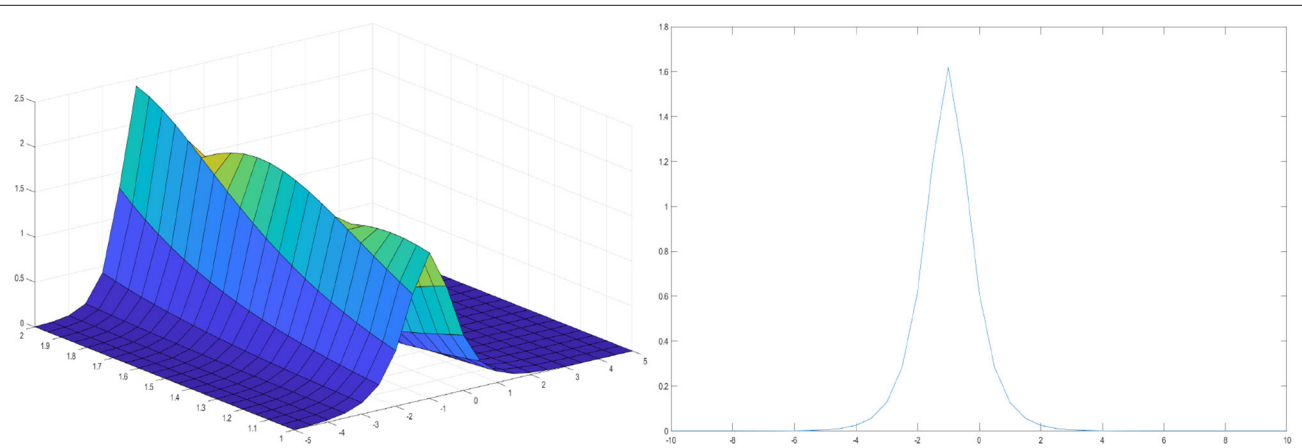


FIGURE 4 | Solitary wave solution of the loaded modified KdV equation for $k = 1$, $\alpha = -0.5$, $p = 1$, and $\gamma(t) = -t$.

7. SOLUTIONS OF THE LOADED GARDNER EQUATION

Assume that Equation (6) has an exact solution in the form of a traveling wave

$$u(x,t) = u(\xi),\ \xi = px - kt,$$

that will convert Equation (6) to an ordinary differential equation

$$-ku' - 2\alpha puu' - 3\beta u^2 u' + p^3 u''' + \gamma(t)pu(0,t)u' = 0. \quad (31)$$

Integrating once Equation (31) with respect to ξ, and putting the constant of integration at zero, we have

$$u'' = \frac{1}{p^3}\left(\beta u^3 + \alpha p u^2 + \left(k - \gamma(t)pu(0,t)\right)u\right). \quad (32)$$

Following Equation (13), it is easy to deduce from Equation (32) an expression for the function $F(u)$

$$\frac{1}{2}\frac{d\left(F^2(u)\right)}{du} = \frac{1}{p^3}\left(\beta u^3 + \alpha p u^2 + \left(k - \gamma(t)pu(0,t)\right)u\right). \quad (33)$$

Integrating Equation (33) and setting the constant of integration to zero yields

$$F(u) = \frac{u}{p}\sqrt{\lambda u^2 + \tau u + \mu(t)}, \quad (34)$$

where $\lambda = \frac{\beta}{2p}, \tau = \frac{2\alpha}{3}, \mu(t) = \frac{k-\gamma(t)pu(0,t)}{p}$. From Equation (12) and Equation (34), we deduce that

$$\frac{du}{u\sqrt{\lambda u^2 + \tau u + \mu(t)}} = \frac{1}{p}d\xi. \quad (35)$$

After integrating Equation (35), with zero constant of integration, we have the following exact solution

$$u(x,t) = \frac{2\mu(t)e^{-\frac{\sqrt{\mu(t)}}{p}(px-kt)}}{\left(e^{-\frac{\sqrt{\mu(t)}}{p}(px-kt)} - \frac{\tau}{2}\right)^2 - \frac{\tau}{2}^2 - \lambda\mu(t)}. \quad (36)$$

It is obvious that the function $u(0,t)$ can be easily found based on expression (36).

We have several types of traveling wave solutions of the loaded Gardner equation as follows:
1) When $\sqrt{\mu(t)} > 0$, we have the periodic wave solution

$$u(x,t) = \frac{2\mu(t)e^{-\frac{\sqrt{\mu(t)}}{p}(px-kt)}}{\left(e^{-\frac{\sqrt{\mu(t)}}{p}(px-kt)} - \frac{\tau}{2}\right)^2 - \frac{\tau}{2}^2 - \lambda\mu(t)}.$$

2) When $\sqrt{\mu(t)} < 0$, we have the solitary wave solution

$$u(x,t) = \frac{2\mu(t)e^{-\frac{\sqrt{\mu(t)}}{p}(px-kt)i}}{\left(e^{-\frac{\sqrt{\mu(t)}}{p}(px-kt)i} - \frac{\tau}{2}\right)^2 - \frac{\tau}{2}^2 - \lambda\mu(t)}.$$

Now, by choosing free parameters, we will write the traveling wave solutions of the loaded Gardner equation in the simple form which can be used for the graphical illustrations.

If $k = 1, \alpha = 3, p = -1, \beta = 2, \gamma(t) = -t$, then we have

$$u(x,t) = \frac{2\left(tu(0,t) - 1\right)e^{\sqrt{tu(0,t)-1}(x+t)}}{\left(e^{\sqrt{tu(0,t)-1}(x+t)} - 1\right)^2 + tu(0,t) - 1}, \quad (37)$$

It is obvious that the function $u(0,t)$ can be easily found based on expression (37).

If $k = 1, \alpha = 3, p = -1, \beta = 2, \gamma(t) = t$, then we have

$$u(x,t) = -\frac{2\left(tu(0,t) + 1\right)e^{\sqrt{tpu(0,t)+1}(x+t)i}}{\left(e^{\sqrt{tpu(0,t)+1}(x+t)i} - 1\right)^2 - \left(tu(0,t) + 1\right)}, \quad (38)$$

It is obvious that the function $u(0,t)$ can be easily found based on expression (38).

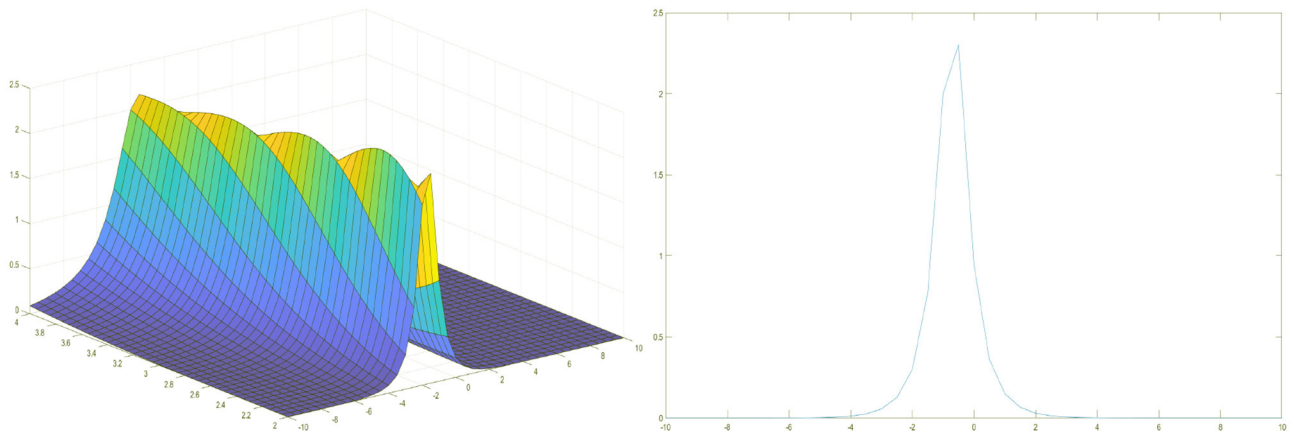


FIGURE 5 | Solitary wave solution of the loaded Gardner equation for $k = 1$, $\alpha = 3$, $p = -1$, $\beta = 2$, $\gamma(t) = -t$.

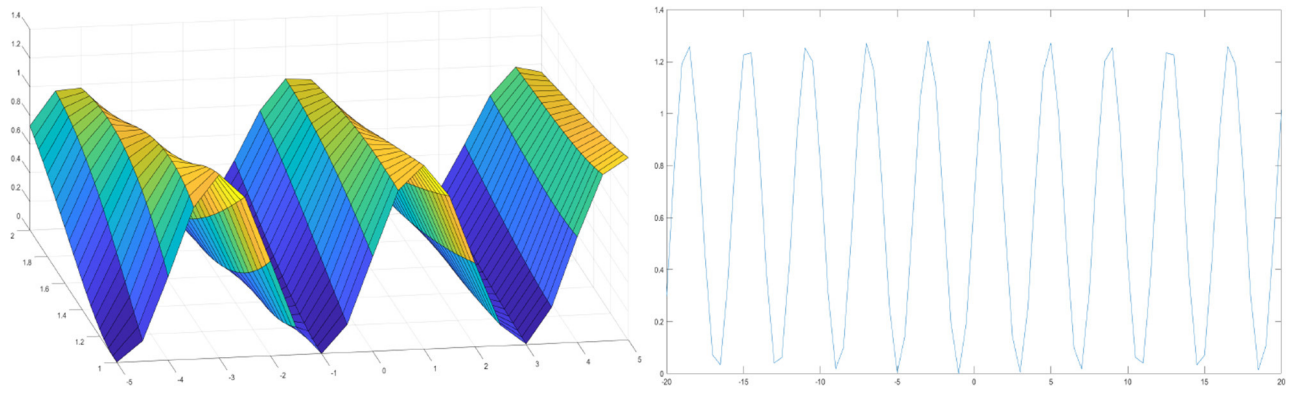


FIGURE 6 | Periodic wave solution of the loaded Gardner equation for $k = 1$, $\alpha = 3$, $p = -1$, $\beta = 2$, $\gamma(t) = t$.

8. GRAPHICAL REPRESENTATIONS OF TRAVELING WAVE SOLUTIONS OF THE LOADED GARDNER EQUATION

This section aims to present graphical illustrations of the obtained traveling wave solutions of the Gardner equation. Using mathematical software Matlab, 3D plots of the obtained solutions have been shown in **Figures 5**, **6**. In the concept of mathematical physics, a soliton or solitary wave is defined as a self-reinforcing wave packet that upholds its shape. At the same time, it propagates at a constant amplitude and velocity. Solitary waves can be obtained from each traveling wave solution by setting particular values to its unknown parameters. By adjusting these parameters, one can get an internal localized mode. We have presented some graphs of solitary waves constructed by taking suitable values of the involved unknown parameters to visualize the underlying mechanism of the original physical phenomena.

9. CONCLUSION

The functional variable method has been successfully used to obtain several traveling wave solutions of the loaded KdV, the loaded modified KdV, and the loaded Gardner equation. The method does not require linearization of differential equations because it is a method of directly solving some non-linear physical models. A wide and general class of modern examples representing real physical problems from plasma physics, fluid dynamics, non-linear optics, and non-linear fields of gas dynamics can be solved easily and elegantly using this method. The exactness of the obtained results is studied by using the software Matlab. The received solutions with free parameters may be important to explain some physical phenomena. The advantage of the method is to give more solution functions such as periodic solutions and hyperbolic solutions than other popular analytical methods. We conclude that the functional variable method is significant and important for finding the exact traveling wave solutions of non-linear evolution equations. The proposed method can be applied to many other non-linear evolution equations in mathematical physics.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

BB and FA conceived of the presented idea. BB developed the theory and performed the computations. FA verified the analytical methods. Both authors discussed the results and contributed to the final manuscript. Both authors contributed to the article and approved the submitted version.

REFERENCES

1. Sagdeev RZ. Cooperative phenomena and shock waves in collision less plasmas. *Rev Plasma Phys*. (1966) 4:23–91.
2. Seadawy AR, Nadia C. Some new families of spiky solitary waves of one-dimensional higher-order KdV equation with power law nonlinearity in plasma physics. *Indian J Phys*. (2020) 94:117–26. doi: 10.1007/s12648-019-01442-6
3. Seadawy AR, Nadia C. Propagation of nonlinear complex waves for the coupled nonlinear Schrödinger Equations in two core optical fibers. *Phys A Stat Mech Appl*. (2019) 529:13–30. doi: 10.1016/j.physa.2019.121330
4. Seadawy AR, Nadia C. Applications of extended modified auxiliary equation mapping method for high order dispersive extended nonlinear Schrödinger equation in nonlinear. *Modern Phys Lett B*. (2019) 33:1–11. doi: 10.1142/S0217984919502038
5. Nadia C, Seadawy AR, Sheng C. More general families of exact solitary wave solutions of the nonlinear Schrödinger equation with their applications in nonlinear optics. *Eur Phys J Plus*. (2018) 133:547. doi: 10.1140/epjp/i2018-12354-9
6. Tagare SG, Singh SV, Reddy RV, Lakhina GS. Electron-acoustic solitons in the Earth's magnetotail. *Nonlinear Process Geophys*. (2004) 11:215–8. doi: 10.5194/npg-11-215-2004
7. Kakad AP, Singh SV, Reddy RV, Lakhina GS, Tagare SG. Electron acoustic solitary waves in the earth's magnetotail region. *Adv Space Res*. (2009) 43:1945–9. doi: 10.1016/j.asr.2009.03.005
8. Abu Arqub O, Rashaideh H. The RKHS method for numerical treatment for integrodifferential algebraic systems of temporal two-point BVPs. *Neural Comput Appl*. (2018) 30:2595–606. doi: 10.1007/s00521-017-2845-7
9. Abu Arqub O. Computational algorithm for solving singular Fredholm time-fractional partial integrodifferential equations with error estimates. *J Appl Math Comput*. (2019) 59:227–43. doi: 10.1007/s12190-018-1176-x
10. Momani S, Abu Arqub O, Maayah B. Piecewise optimal fractional reproducing Kernel solution and convergence analysis for the Atangana-Baleanu-Caputo model of the Lienard's equation. *Fractals*. (2020) 28:2040007. doi: 10.1142/S0218348X20400071
11. Momani S, Maayah B, Abu Arqub O. The reproducing kernel algorithm for numerical solution of Van der Pol damping mod/el in view of the Atangana-Baleanu fractional approach. *Fractals*. (2020) 28:2040010. doi: 10.1142/S0218348X20400101
12. Gorbacheva OB, Ostrovsky LA. Nonlinear vector waves in a mechanical model of a molecular chain. *Phys D Nonlinear Phenomena*. (1983) 8:223–8. doi: 10.1016/0167-2789(83)90319-6
13. Erbay S, Suhubi ES. Nonlinear wave propagation in micropolar media. II: Special cases, solitary waves and Painleve analysis. *Int J Eng Sci*. (1989) 27:915–9. doi: 10.1016/0020-7225(89)90032-3
14. Zha QL, Li ZB. Darboux transformation and multi-solitons for complex mKdV equation. *Chin Phys Lett*. (2008) 25:8. doi: 10.1088/0256-307X/25/1/003
15. Karney CFF, Sen A, Chu FYF. Nonlinear evolution of lower hybrid waves. *Phys Fluids*. (1979) 22:940–52. doi: 10.1063/1.862688
16. Hamdi S, Morse B, Halphen B, Schiesser W. Analytical solutions of long nonlinear internal waves: part I. *Nat Hazards*. (2011) 57:597–607. doi: 10.1007/s11069-011-9757-0
17. Naher H, Aini Abdullah F, Ali Akbar M. New traveling wave solutions of the higher dimensional nonlinear partial differential equation by the exp-function method. *J Appl Math*. (2012) 2012:575387. doi: 10.1155/2012/575387
18. Akbar MA, Norhashidah MA. Exp-function method for Duffing equation and new solutions of (2+1) dimensional dispersive long wave equations. *Prog Appl Math*. (2011) 1:30–42. doi: 10.3968/j.pam.1925252820120102.003
19. Baldwin D, Göktas Ü, Hereman W, Hong L, Martino RS, Miller JC. Symbolic computation of exact solutions expressible in hyperbolic and

elliptic functions for nonlinear PDEs. *J Symb Comput*. (2004) 37:669–705. doi: 10.1016/j.jsc.2003.09.004
20. Hereman W, Nuseir A. Symbolic methods to construct exact solutions of nonlinear partial differential equations. *Math Comput Simul*. (1997) 43:13–27. doi: 10.1016/S0378-4754(96)00053-5
21. Demiray H. Variable coefficient modified KdV equation in fluid-filled elastic tubes with stenosis: solitary waves. *Chaos Solitons Fractals*. (2009) 42:358–64. doi: 10.1016/j.chaos.2008.12.014
22. Kudryashov NA. Exact solutions of the generalized Kuram-Oto-Sivashinsky equation. *Phys Lett*. (1990) 24:287–91. doi: 10.1016/0375-9601(90)90449-X
23. Chen Y, Wang Q. Extended Jacobi elliptic function rational expansion method and abundant families of Jacobi elliptic functions solutions to (1+1)-dimensional dispersive long wave equation. *Chaos Solitons Fractals*. (2005) 24:745–57. doi: 10.1016/j.chaos.2004.09.014
24. Malfliet W. Solitary wave solutions of nonlinear wave equations. *Am J Phys*. (1992) 60:650–4. doi: 10.1119/1.17120
25. Ablowitz MJ, Clarkson PA. *Solitons, Nonlinear Evolution Equations and Inverse Scattering Transform*. New York: Cambridge University Press (1991). doi: 10.1017/CBO9780511623998
26. Hirota R. Exact solution of the KdV equation for multiple collisions of solutions. *Phys Rev Lett*. (1971) 27:1192–4. doi: 10.1103/PhysRevLett.27.1192
27. Rogers C, Shadwick WF. *Backlund Transformations and Their Applications*. New York, NY: Academic Press (1982).
28. He JH, Wu XH. Exp-function method for nonlinear wave equations. *Chaos Solitons Fractals*. (2006) 30:700–8. doi: 10.1016/j.chaos.2006.03.020
29. Kudryashov NA. On types of nonlinear non-integrable equations with exact solutions. *Phys Lett A*. (1991) 155:269–75. doi: 10.1016/0375-9601(91)90481-M
30. Abdou MA, Soliman AA. Modified extended tanH-function method and its application on nonlinear physical equations. *Phys Lett A*. (2006) 353:487–92. doi: 10.1016/j.physleta.2006.01.013
31. Zhao X, Tang D. A new note on a homogeneous balance method. *Phys Lett A*. (2002) 297:59–67. doi: 10.1016/S0375-9601(02)00377-8
32. Kneser A. Belastete integralgleichungen. *Rendiconti del Circolo Matematico di Palermo*. (1914) 37:169–97. doi: 10.1007/BF03014816
33. Lichtenstein L. *Vorlesungen über einige Klassen nichtlinearer Integralgleichungen und Integro-Differential-Gleichungen nebst Anwendungen*. Berlin: Springer-Verlag (1931). doi: 10.1007/978-3-642-47600-6
34. Nakhushev AM. *Equations of Mathematical Biology*. Moscow: Visshaya Shkola (1995).
35. Nakhushev AM. Loaded equations and their applications. *Diff Equat*. (1983) 19:86–94.
36. Nakhushev AM. The Darboux problem for a certain degenerate second order loaded integrodifferential equation. *Diff Equat*. (1976) 12:103–8.
37. Nakhushev AM, Borisov VN. Boundary value problems for loaded parabolic equations and their applications to the prediction of ground water level. *Diff Equat*. (1977) 13:105–10.
38. Nakhushev AM. Boundary value problems for loaded integro-differential equations of hyperbolic type and some of their applications to the prediction of ground moisture. *Diff Equat*. (1979) 15:96–105.
39. Baltaeva UI. On some boundary value problems for a third order loaded integro-differential equation with real parameters. *Bull Udmurt Univ Math Mech Comput Sci*. (2012) 3:3–12. doi: 10.20537/vm120301
40. Kozhanov AI. A nonlinear loaded parabolic equation and a related inverse problem. *Math Notes*. (2004) 76:784–95. doi: 10.1023/B:MATN.0000049678.16540.a5
41. Hasanov AB, Hoitmetov UA. Integration of the general loaded Kortewegde Vries equation with an integral source in the class of rapidly decreasing complex-valued functions. *Russian Math*. (2021) 7:52–66. doi: 10.3103/S1066369X21070069
42. Hasanov AB, Hoitmetov UA. *On Integration of the Loaded Korteweg-de Vries Equation in the Class of Rapidly Decreasing Functions*. Proceedings of the Institute of Mathematics and Mechanics. National Academy of Sciences of Azerbaijan. (2021). p. 250–61. doi: 10.30546/2409-4994.47.2.250
43. Khasanov AB, Hoitmetov UA. On integration of the loaded mKdV equation in the class of rapidly decreasing functions. *Bull Irkutsk State Univ Ser Math*. (2021) 38:19–35. doi: 10.26516/1997-7670.2021.38.19
44. Urazboev GU, Baltaeva II, Rakhimov ID. Generalized G'/G - extension method for loaded Korteweg-de Vries equation. *Siberian J Indus Math*. (2021) 24:72–6. doi: 10.33048/sibjim.2021.24.410
45. Yakhshimuratov AB, Matyokubov MM. Integration of a loaded Kortewegde Vries equation in a class of periodic functions. *Russian Math*. (2016) 60:72–6. doi: 10.3103/S1066369X16020110
46. Zerarka A, Ouamane S, Attaf A. On the functional variable method for finding exact solutions to a class of wave equations. *Appl Math Comput*. (2010) 217:2897–904. doi: 10.1016/j.amc.2010.08.070
47. Zerarka A, Ouamane S. Application of the functional variable method to a class of nonlinear wave equations. *World J Modell Simul*. (2010) 6:150–60.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Applied Mathematics and Statistics

ORIGINAL RESEARCH
published: 30 June 2022
doi: 10.3389/fams.2022.940133

Early Warning Signals of Financial Crises Using Persistent Homology and Critical Slowing Down: Evidence From Different Correlation Tests

Mohd Sabri Ismail, Mohd Salmi Md Noorani, Munira Ismail and Fatimah Abdul Razak*

Department of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Malaysia

In this study, a new market representation from persistence homology, known as the L^1-norm time series, is used and applied independently with three critical slowing down indicators [autocorrelation function at lag 1, variance, and mean for power spectrum (MPS)] to examine two historical financial crises (Dotcom crash and Lehman Brothers bankruptcy) in the US market. The captured signal is the rising trend in the indicator time series, which can be determined by Kendall's tau correlation test. Furthermore, we examined Pearson's and Spearman's rho correlation tests as potential substitutes for Kendall's tau correlation. After that, we determined a correlation threshold and predicted the whole available date. The point of comparison between these correlation tests is to determine which test is significant and consistent in classifying the rising trend. The results of such a comparison will suggest the best test that can classify the observed rising trend and detect early warning signals (EWSs) of impending financial crises. Our outcome shows that the L^1-norm time series is more likely to increase before the two financial crises. Kendall's tau, Pearson's, and Spearman's rho correlation tests consistently indicate a significant rising trend in the MPS time series before the two financial crises. Based on the two evaluation scores (the probability of successful anticipation and probability of erroneous anticipation), by using the L^1-norm time series with MPS, our result in the whole prediction demonstrated that Spearman's rho correlation (46.15 and 53.85%) obtains the best score as compared to Kendall's tau (42.31 and 57.69%) and Pearson's (40 and 60%) correlations. Therefore, by using Spearman's rho correlation test, L^1-norm time series with MPS is shown to be a better way to detect EWSs of US financial crises.

Keywords: topological data analysis, persistent homology, critical transition, critical slowing down, correlation tests, early warning signal, financial crises, complex system

OPEN ACCESS

Edited by:
Mohd Hafiz Mohd, Universiti Sains Malaysia (USM), Malaysia

Reviewed by:
Majid Khan, Universiti Sains Malaysia (USM), Malaysia
Marian Gidea, Yeshiva University, United States

****Correspondence:***
Mohd Sabri Ismail
sabrimohd92@gmail.com

Specialty section:
This article was submitted to Mathematics of Computation and Data Science, a section of the journal Frontiers in Applied Mathematics and Statistics

Received: *10 May 2022*
Accepted: *07 June 2022*
Published: *30 June 2022*

Citation:
Ismail MS, Md Noorani MS, Ismail M and Abdul Razak F (2022) Early Warning Signals of Financial Crises Using Persistent Homology and Critical Slowing Down: Evidence From Different Correlation Tests. Front. Appl. Math. Stat. 8:940133. doi: 10.3389/fams.2022.940133

INTRODUCTION

Understanding behaviors of financial crises (unexpected and huge declines in the stock market) are crucial to explaining the dynamics of the financial market. However, such financial events are very challenging to study. Among them, the challenges are understanding how the financial market behaves before financial crises and developing a method that is capable to detect early warning signals (EWSs) of the financial crises. The detection method will become more beneficial if it can

accurately distinguish the financial market into two classification periods, which are EWS periods (indicating a possible huge downtrend is coming in the market) and safe periods. Such a method helps investors and traders to develop an early precaution strategy and protect their investments from any losses due to a financial crisis.

In practice, the Chicago Board Options Exchange's CBOE Volatility Index or VIX index—a real-time market index representing the market's expectations for volatility over the coming 30 days—is always used by market participants to alert any upcoming financial crisis. The uptrend patterns in the VIX were observed before financial crises, unfortunately, low-level VIX also reported occurred before financial crises [1, 2]. This suggests that the VIX can provide EWSs; however, a reliable method to predict financial crises remains a challenge in this field. Therefore, extensive studies have been conducted by many researchers in an attempt to provide other rationales to explain why financial crises occurred and provide any other possible EWS. Some of those methods are bubble theory [3–5], financial stress indicator [6, 7], information-based measures [8, 9], financial network analysis [10–12], and graphical analysis [13].

However, this study is focusing on critical transition theory. This theory viewed the financial market as a complex system with episodes of critical transitions (abrupt shifts from a current stable state to another stable state when the system reaches critical points) [14, 15]. When approaching a critical point, this theory stated that a generic phenomenon happened known as critical slowing down (CSD) because of decreasing stability in the market and its recovery rate took longer to preserve the stability. At a critical point, the financial market loses its stability in the dynamic state and then suddenly causes a market movement into a financial crisis. CSD gives a rising trend in the time series of some indicators, such as autocorrelation function at lag 1 (AC1), variance (VAR), and mean for power spectrum (MPS) at low frequencies [16–18].

In the traditional method, the rising trend in the time series of CSD indicator that happens before the critical transition point is determined using Kendall's tau correlation test. By using Kendall's tau correlation test, previous studies have shown that the observed rising trend can provide EWSs before financial crises [19–23]. Despite the successful results obtained, there is a realization that some indicator like AC1 also tends to decline before financial crises. All of these lead to mixing results such as recorded in Guttal et al. [21] and Diks et al. [23]. However, as compared to critical transitions, Guttal et al. [21] argued that financial crises are more likely to follow stochastic transition, in which variability indicators (VAR and MPS) can signal early warnings for financial markets. This also points out that abrupt transitions in the financial market are hardly the same as a complex system in nature such as the earth's global climate and interaction between species in ecology. One of the reasons behind this is the financial market is involved with human behaviors that influence the market's movement [23].

Recently, using persistent homology (PH) (a robust method to compute topological features of financial data at different spatial scales [24]), Gidea and Katz [25] suggested a new market representative obtained from persistence landscapes called L^1-norm time series. Since the persistence landscapes are robust under perturbations of the underlying data, the L^1-normtime series has the advantage to reflect the loss of stability in dynamic states of the original system. When the stock market becomes more volatile, loops in the relevant point clouds become much more pronounced and give more significant features within persistence landscapes [25]. At that time, the corresponding L^1-normvalues are more likely to jump up, and this behavior is believed to correspond with an undergoing CSD. Therefore, to test the presence of CSD, the CSD indicators are used with the L^1-normtime series, where these indicators should alert upcoming financial crises by showing their significant rising trend that can be classified by Kendall's tau correlation test. In their work, Gidea and Katz [25] showed that theL^1-normtime series grew substantially before the Dotcom crash and Lehman Brothers bankruptcy in the US market. Interestingly, by using CSD indicators and Kendall's tau correlation test on the L^1-normtime series, Gidea and Katz [25] confirmed that a significant strong rising trend happen in the MPS time series before those two financial crises. Therefore, the method is suggested as a new potential EWS. Later, the application of PH to financial data analysis has attracted more attention from researchers, especially in detecting EWSs. All the articles published in this area are discussed in Section Literature Review.

In this article, we also used PH and CSD indicators (AC1, VAR, and MPS) to detect EWSs of financial crises. Unlike Gidea and Katz [25], instead of using Kendall's tau correlation test to indicate the rising trend in the time series of the indicators, we also examined Pearson's and Spearman's rho correlation tests as a potential substitute for Kendall's tau correlation. The point of comparison between these correlation tests is to determine which test is significant and consistent in classifying the rising trend. The results of such a comparison will suggest the best test that can classify the observed rising trend and detect EWSs of impending financial crises. The remaining portion of this article is organized as follows. Section Literature Review briefly discusses our literature review, Section Persistent Homology introduces PH, Section Data Analysis elaborates on our data, Section Methods presents the applied methods, Section Result mentions our results and their corresponding discussions, and Section Conclusion wraps our conclusion.

LITERATURE REVIEW

Topological information from financial data obtained using PH has been used to study financial problems. In recent times, this topic has attracted attention from many researchers around the globe. In the finance market, the most application currently investigated by PH is financial crises and their EWS detection tools. In a multivariate setting, by begin with examining chaotic time series with noise and a growing variance, Gidea and Katz [25] showed that PH can exhibit strong growth in its market representative time series prior to the critical point. This growth can be analyzed using CSD indicators (AC1, VAR, and MPS) and correlation tests (Kendall's tau) to indicate corresponding

CSD that happen before critical point. After that, Gidea and Katz [25] used the pipeline method to examine the Dotcom crash and Lehman Brothers bankruptcy in the US market as mentioned earlier.

After that, Ismail et al. [26] expanded the study of Gidea and Katz [25] by using PH and CSD indicators to examine financial crises in the US, Singapore, and Malaysia markets. Aspects of the method's robustness and prediction performance have been rigorously evaluated in Ismail et al. [26]. Meanwhile, Aromi et al. [27] substantiated that PH reflects changes in the underlying multivariate distribution and strong covariance could nullify the persistence of homologies. On the other hand, [28] also analyzed time-dependent correlation networks using PH to detect early signs of critical transitions in financial data.

In Guo et al. [29], an EWS based on PH is also built to detect the critical dates on the financial time series. Guo et al. [30] also used topological features of complex networks that were extracted using PH to find critical dates for financial crises. In addition, dynamics of financial market correlations based on topology and geometry are examined using PH in Yen et al. [31]. Moreover, Yen and Cheong [32]also tested PH to analyze Singapore and Taiwan markets. The extreme event called flash crash was also explored in Kim et al. [33] by applying PH and dynamic time series analysis.

Furthermore, anomalies detection in the dynamics of a market index also is studied with PH [34]. Katz and Biem [35] also showed that early signatures of growing market instabilities can be captured by PH. Besides, Gidea et al. [3] used PH and k-means clustering to detect critical transitions in the time series of cryptocurrencies. For Bitcoin [36], also uses PH and CSD indicators to detect such transitions and substantiated that PH can detect EWSs better than the detrending time-series approach (the most common approach used in previous studies to detect CSD indicators).

Other than that, PH also has been developed to improve portfolio investment strategies. Studied ten global indices and all their underlying assets, Goel et al. [37] showed that a new strategy based on PH leads to more robust portfolios. Baitinger and Flegel [38] also demonstrated that investment strategies relying on a PH-based turbulence detection outperform investment strategies based on other popular turbulence indices. In clustering and classification of financial time series, Majumdar and Laha [39] showed that PH outperforms other methods in this task.

Additionally, PH is applied with machine learning to predict the movement of financial data. Such a task has been done in Ismail et al. [40] by using PH and machine learning methods (logistic regression, neural network, support vector machine, and random forest) to predict the next-day direction of the Kuala Lumpur Composite Index (KLCI). Moreover, Baitinger and Flegel [41] also introduced PH to produce microstructural predictors, where these predictors are combined with machine learning and statistical factor extraction methods to predict asset returns.

PERSISTENT HOMOLOGY

Persistent homology is a new quantitative method of topological data analysis to compute topology features (connected components, loops, voids, and others) that persistently emerge across multiple scales. Interestingly, PH is robust to small perturbations of input data, independent of dimensions and coordinates and provides a compact representation of the qualitative features of the input [42–44]. These characteristics make PH suitable to analyze complex, non-linear, noisy, and high-dimensional data like financial data [45].

Later, a summary of PH in the way that it is used in this article is provided. We noted that the concepts of PH stated here can also be found in most books and journals related to PH. To further explore the theories and other concepts of PH, we recommend Otter et al. [24] and Edelsbrunner and Harer [46] to the interested readers.

Input data analyzed by PH is called point cloud dataset (PCD), which can be denoted as $X = \left\{x_i \in \mathbb{R}^d | i = 1, \cdots, n\right\}$ for $d \geq 2$. Let X be a PCD, a Rips complex at a scale $\varepsilon > 0$ (denoted as $R(X, \varepsilon)$) can be constructed as follows:

- For each dimensional $k = 0, 1, \cdots$, a k-simplex of X points $\left\{x_{i_1}, \cdots, x_{i_{(k+1)}}\right\}$ belongs to a $R(X, \varepsilon)$ if and only if for every pair $\left\{x_{i_r}, x_{i_s}\right\}$, we have $\left|x_{i_r} - x_{i_s}\right| \leq \varepsilon$ for all $x_{i_r}, x_{i_s} \in \left\{x_{i_1}, \cdots, x_{i_{(k+1)}}\right\}$ [24].

Roughly speaking, a Rips complex is a combination of vertices (0-simplex), edges (1-simplex), solid triangles (2-simplex), and higher dimensional analog, joined according to the above-mentioned rule. **Figure 1** illustrates 0-simplex until 2-simplex and a two-dimensional Rips complex $R(X, \varepsilon)$,respectively. In **Figure 1**, the scale ε is the diameter of balls around each point in the different PCDs.

In a real-world application, finding a single ε is impractical since the real space behind PCD is mostly unknown. Therefore, PH provides a better way to interpret topological information regarding the shape behind PCD by varying the scale ε. Let $\varepsilon \in \{\varepsilon_0, \cdots \varepsilon_m\}$ such that $0 \leq \varepsilon_0 < \cdots < \varepsilon_m$, then filtration of Rips complexes is $R(X, \varepsilon_0) \subset \cdots \subset R(X, \varepsilon_m)$. **Figure 2** shows a filtration containing six Rips complexes constructed of a PCD at multiple scales.

For $i = 1, \cdots, m$, we can compute the k-homology of each Rips complex $R(X, \varepsilon_i)$, denoted as $H_k(R(X, \varepsilon_i))$. Roughly speaking, $H_0(R(X, \varepsilon_i))$ is the free group generated by the connected components of $R(X, \varepsilon_i)$, $H_1(R(X, \varepsilon_i))$ is the free group generated by the loops in $R(X, \varepsilon_i)$, $H_2(R(X, \varepsilon_i))$ is the free group generated by the voids of $R(X, \varepsilon_i)$. The Betti numbers count the number of generators of such homology groups. This number of generators indicates the number of corresponding topological features that emerge at each scale.

In addition, information regarding the lifespans of topological features also can be obtained by using PH. In the PH community, most believe that a topological feature that has a longer lifespan (persists for a bigger range of scales) can be viewed as a more

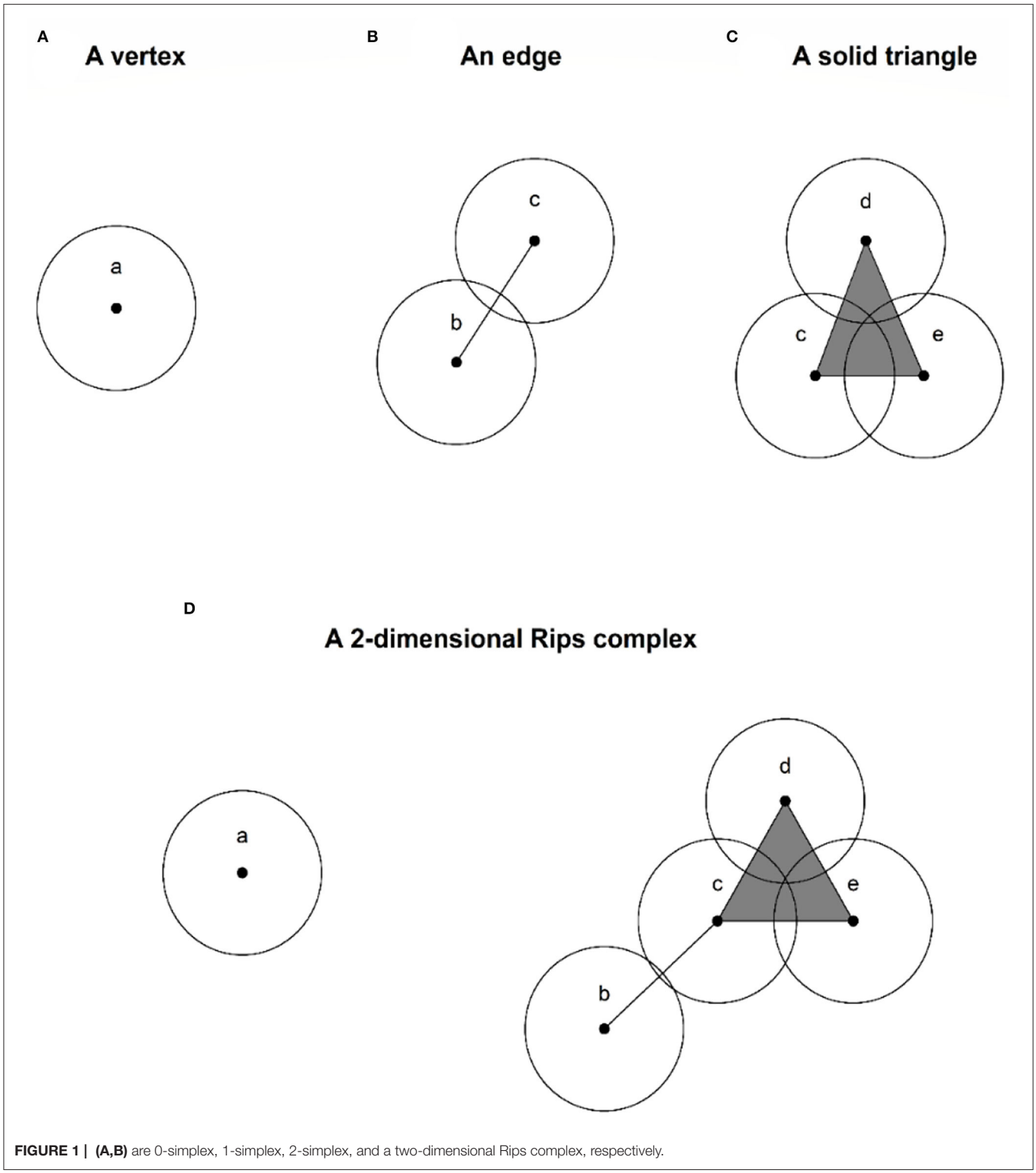


FIGURE 1 | (A,B) are 0-simplex, 1-simplex, 2-simplex, and a two-dimensional Rips complex, respectively.

significant one, whereas a feature that has a shorter lifespan (persists for a smaller range) can be viewed as a less significant, or a noisy feature. Nevertheless, the theoretical justification for this is unclear and may be dependent on the problem at hand [45]. However, this study used all the obtained lifespan in our computation.

Let $\left\{(\varepsilon_b, \varepsilon_d)_i | i = 1, \cdots, n\right\}$ be a collection of the lifespans of topological features such that $\varepsilon_b < \varepsilon_d$ and $\varepsilon_b, \varepsilon_d \in \{\varepsilon_0, \cdots \varepsilon_m\}$.

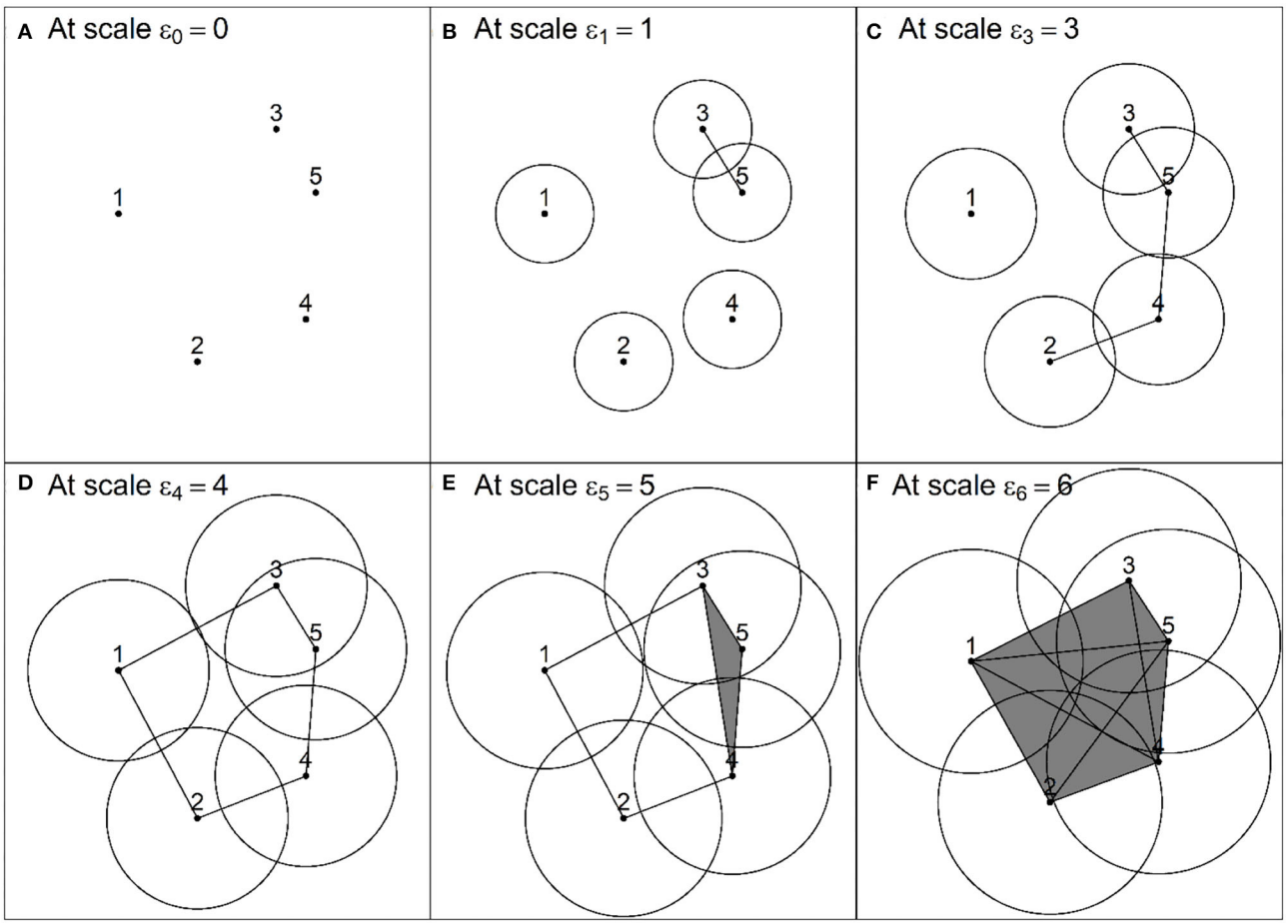


FIGURE 2 | Filtration of six Rips complexes constructed of a PCD at six different scales.

The simplest tool to present this information is a persistent barcode, which is a collection of n half-closed intervals $[\varepsilon_b, \varepsilon_d)$ representing the lifespans of topological features, see **Figure 3A**. Other tools are a persistent diagram and a persistent landscape. A persistent diagram is a finite collection of n birth–death points $(\varepsilon_b, \varepsilon_d) \in \mathbb{R}^2$ that lie along or above a diagonal line. If there are redundant birth–death points, then the points will be represented as a single point but multiplied by its size to correspond to the frequency of this point. **Figure 3B** provides an example of a persistent diagram.

On the other hand, the persistence landscape λ is quite a recent tool introduced by Bubenik and Dłotko [47] and Bubenik [48]. To define the persistent landscape, we transformed each lifespan into a piecewise linear function $f_{(\varepsilon_b,\varepsilon_d)_i} : \rightarrow [0, \infty)$, which is defined below:

$$f_{(\varepsilon_b,\varepsilon_d)_i}(x) = \begin{cases} x - \varepsilon_b & \text{if } x \in (\varepsilon_b, (\varepsilon_b + \varepsilon_d)/2\,] \\ -x + \varepsilon_d & \text{if } x \in ((\varepsilon_b + \varepsilon_d)/2\,, \varepsilon_d) \\ 0 & \text{if } x \notin ((\varepsilon_b + \varepsilon_d)) \end{cases},$$
$$\text{for } i = 1, \cdots, n. \tag{1}$$

As a result, we obtained a set of n piecewise linear functions $\left\{f_{(\varepsilon_b,\varepsilon_d)_i}(x) \,|i = 1, \cdots, n\right\}$. Later the layer in the persistence landscape was obtained for each $k \in \mathbb{Z}$, which is a function $\lambda_k(x) : \mathbb{R} \rightarrow [0, \infty)$that is defined as follows $\lambda_k(x) = k - \max\left(\left\{f_{(\varepsilon_b,\varepsilon_d)_i}(x) \,|i = 1, \cdots, n\right\}\right)$, where $k - \max$ denotes the $k$thlargest value of the piecewise linear functions. If the kth largest value does not exist anymore for $k = l$, then $\lambda_k(x) = 0$for all remaining $k \geq l$. The persistent landscape λ is the infinite sequence of $\lambda_k(x)$, which can be denoted as $\{\lambda_1(x), \lambda_2(x), \cdots\}$. **Figures 3C,D** presents the persistence landscapes of connected components and a loop, respectively.

In this study, we used persistent landscapes because this representation can be summarized into a summary point, that is a L^p-norm. This norm is a function $\|.\|_p : \lambda \rightarrow \mathbb{R}$, which is defined as follows:

$$\|\lambda\|_p = \left[\sum_{k=1}^{\infty} \int |\lambda_k(x)|^p dx\right]^{\frac{1}{p}} \text{ for } 1 \leq p < \infty, \tag{2}$$
$$\text{and } \|\lambda\|_\infty = \sup_{k,x} |\lambda_k(x)| \text{ for } p = \infty. \tag{3}$$

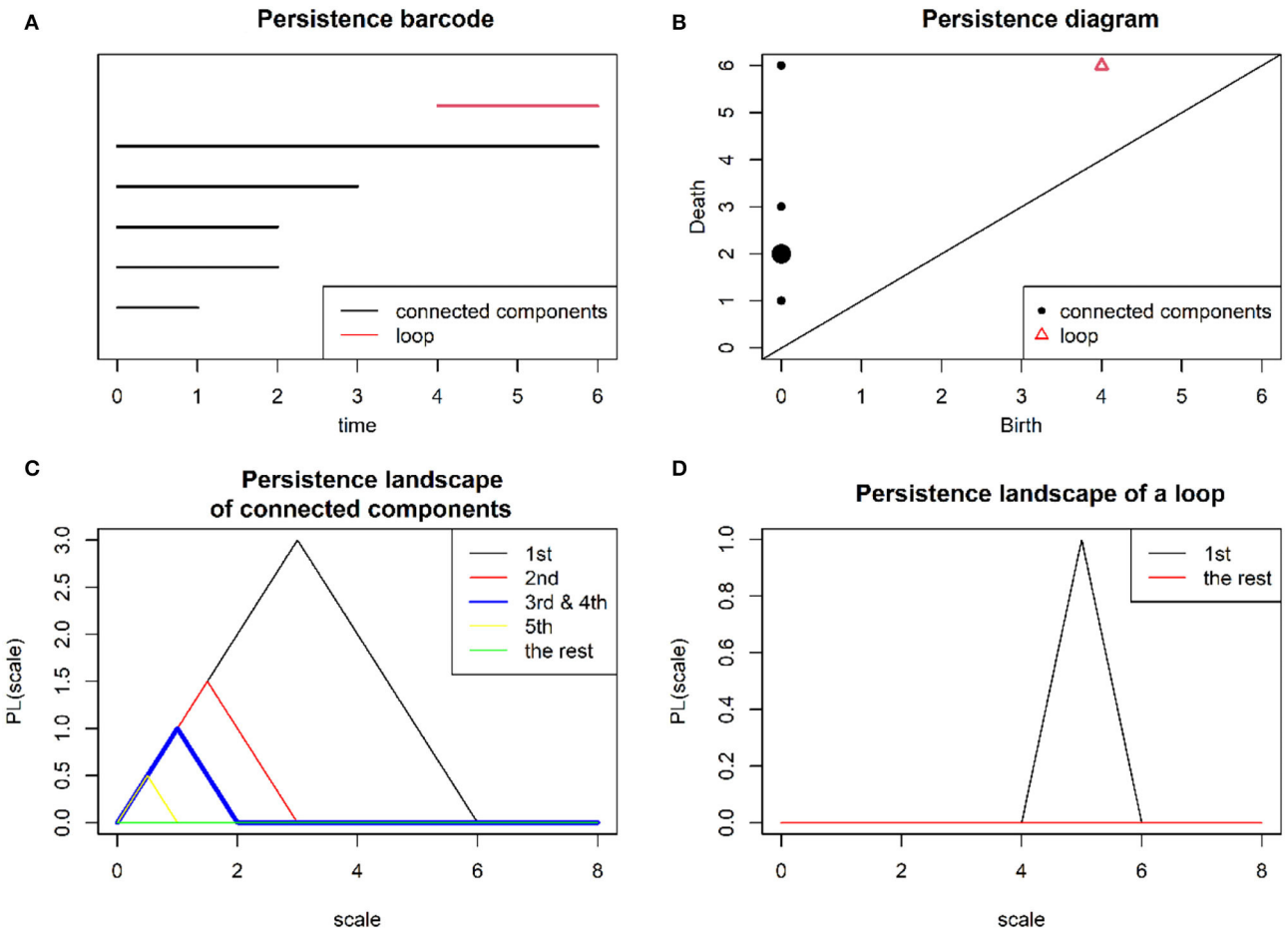


FIGURE 3 | Refer to the filtration of Rips complexes in **Figure 2**. By applying PH, compact representations: **(A)** a persistent barcode, **(B)** a persistence diagram (note that the size for point (0, 2) is enlarged two times bigger than the others to indicate two connected components that share a redundant lifetime), **(C)** a persistence landscape of connected components (note that the third sub-landscape is enlarged two times bigger than the others to indicate two connected components, which share a redundant lifetime), and **(D)** a persistence landscape of a loop are obtained.

Furthermore, statistical properties of these norm values also can be analyzed. Therefore, the CSD indicators are computed from these norm values, and they are used to detect EWSs of impending financial crises in this study.

SAMPLE DATA

For this study, four main stock indices of the US market were collected, which span from 22/12/1987 until 29/12/2017. The indices were the Standard and Poor's 500 (S&P 500), the Dow Jones Industrial Average (DJIA), the Nasdaq Composite (Nasdaq), and the Russell 2000 Index (Russell 2000), which were derived from the Yahoo Finance. All these indices are shown in **Figure 4**. In **Figure 4**, for illustration purposes, the indices are normalized using the max/min normalization using the formula $p_{norm} = 2 \times \left(\left(p_t - \min\right) / \left(\max - \min\right)\right)$, where p_t,max,and min are the closing price at date t, the maximum closing price, and the minimum closing price of an index, respectively. Such normalization does not involve our method as discussed in Section Methods. Furthermore, in **Figure 4**, the crisis dates for the Dotcom crash and Lehman Brothers bankruptcy are mentioned.

METHODS

Pre-crisis Dataset

In this study, two pre-crisis datasets were generated. The two datasets contained closing prices of the four indices (S&P 500, DJIA, Nasdaq, and Russell 2000) with a length of 1000 before the crisis date of the Dotcom crash and Lehman Brothers bankruptcy, respectively. The objective of creating the two pre-crisis datasets was to study the financial market's behaviors before the financial crises, where the obtained information based on PH and CSD would be proceeded to form a EWS detection. Later, we predicted and compared the performance for the case of using different correlation tests: Kendall's tau, Pearson's, and Spearman's rho.

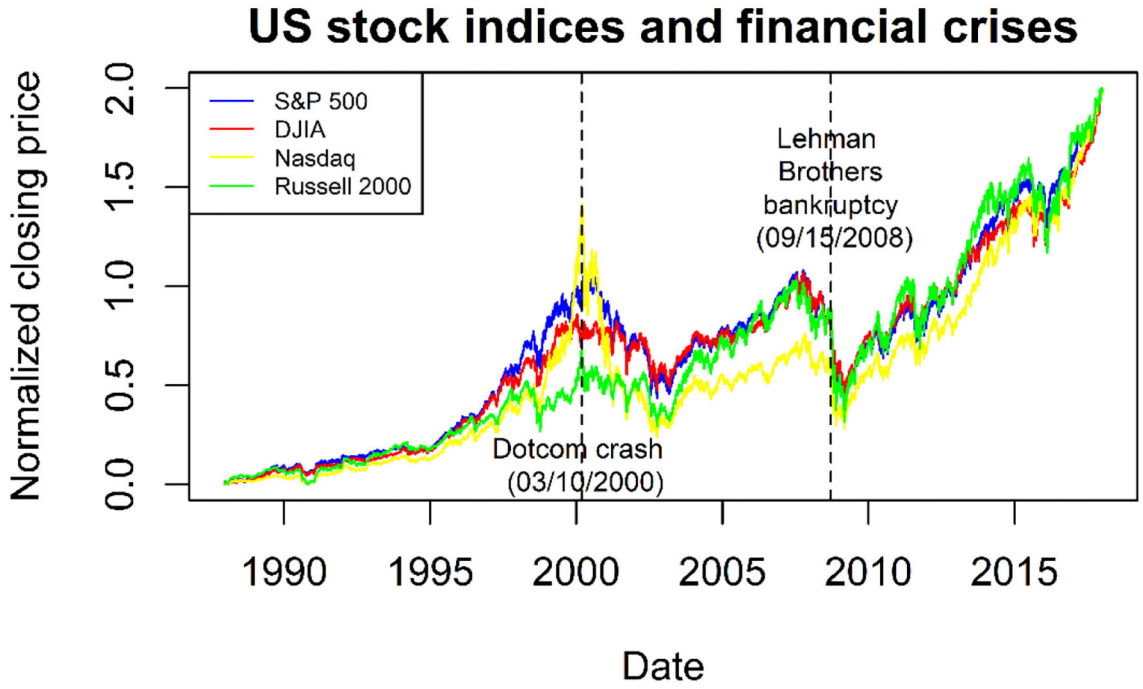


FIGURE 4 | Indices for the Standard and Poor's 500 (S&P 500), the Dow Jones Industrial Average (DJIA), the Nasdaq Composite (Nasdaq), and the Russell 2000 Index (Russell 2000), respectively. The crisis dates for the Dotcom crash and Lehman Brothers bankruptcy also are mentioned.

Persistent Homology

For $i = 1, \cdots, 4$, 1000 closing prices of an index i in the pre-crisis were transformed to daily log returns using the formula $r_t^i = \ln\left(p_t^i / p_{t-1}^i\right)$ for $t = 1, \cdots, 1000$, where p_t^i is the closing price at date t. As a result, for each pre-crisis dataset, we obtained a transformed pre-crisis dataset, denoted as $X = \left\{r_t \in \mathbb{R}^4 | t = 1, \cdots, 1000\right\}$, where $r_t = \left(r_t^1, \cdots, r_t^4\right)$ is a point of the X at the date t. Xcan be illustrated in matrix form as below:

$$X = \begin{bmatrix} r_1^1 & r_1^2 & \cdots & r_1^4 \\ r_2^1 & r_2^2 & \cdots & r_2^4 \\ \vdots & \vdots & \ddots & \vdots \\ r_{1000}^1 & r_{1000}^2 & \cdots & r_{1000}^4 \end{bmatrix} \tag{4}$$

Furthermore, the daily sliding window of length 50 approach was applied to segment the X to obtain PCDs. For financial data analysis, it had been demonstrated that a length of 50 was enough to extract topological information through PH as reported in the previous literature reviews [3, 25, 26, 35]. Consequently, each respective PCD from X at the date t also can be illustrated in matrix form as follows:

$$X(t) = \begin{bmatrix} r_{((t-50)+1)}^1 & r_{((t-50)+1)}^2 & \cdots & r_{((t-50)+1)}^4 \\ r_{((t-50)+2)}^1 & r_{((t-50)+2)}^2 & \cdots & r_{((t-50)+2)}^4 \\ \vdots & \vdots & \ddots & \vdots \\ r_{((t-50)+50)}^1 & r_{((t-50)+50)}^2 & \cdots & r_{((t-50)+50)}^4 \end{bmatrix}, \quad \text{for } t = 1, \cdots, 1000. \tag{5}$$

For each t, as briefed in Section Persistent Homology, PH of Rips filtration that built on PCD $X(t)$,the corresponding persistence landscape at the date t and the corresponding L^1-norm value at the date t were computed, accordingly. By doing so, we obtained L^1-normtime series, denoted by $\mathrm{Y} = \left\{\|\lambda\|_{1,t} | t = 50, \cdots, 1000\right\}$, where $\|\lambda\|_{1,t}$ is a L^1-norm value at the date t.

Critical Slowing Down Indicators

Consequently, the L^1-normtime series $\mathrm{Y} = \left\{\|\lambda\|_{1,t} | t = 50, \cdots, 1000\right\}$ was segmented by using the daily sliding window of the length 500 to obtain sequences of 500 L^1-norm values. The length 500 was selected based on the literature mentioned in Guttal et al. [21], Diks et al.[23], and Ismail et al. [26], which uses half of the length of the pre-crisis dataset. Each sequence of 500 L^1-norm values at the date t can be denoted as below:

$$\mathrm{Y}(t) = \left\{\|\lambda\|_{1,((t-500)+1)}, \cdots, \|\lambda\|_{1,((t-500)+500)}\right\}, \quad \text{for } t = 549, \cdots, 1000. \tag{6}$$

For each $t \in \{549, \cdots, 1000\}$, we computed a value based on the CSD indicators: AC1, VAR, and MPS at low frequencies as accordingly defined below:

- The AC1 value at trading t is $ac1_t = {}^{\rho_{1,t}}\!/_{\mathrm{var}\ t}$, where $\rho_{1,t} = \frac{1}{500-1} \sum_{j=((t-500)+1)}^{((t-500)+500)-1} \left(\|\lambda\|_{1,j} - \overline{\|\lambda\|}_{1,j}\right)\left(\|\lambda\|_{1,(j+1)} - \overline{\|\lambda\|}_{1,j}\right)$, [[Mathtype-mtef1-eqn-106.mtf]] is the mean of $\mathrm{Y}(t)$ and var_t is the VAR of $\mathrm{Y}(t)$ as defined in the point below.

- The VAR value at the date t is $\text{var}_t = \frac{1}{500-1}\sum_{j=((t-500)+1)}^{((t-500)+500)-1}\left(\|\lambda\|_{1,j} - \overline{\|\lambda\|}_{1,j}\right)^2$, where $\overline{\|\lambda\|}_{1,j} = \frac{1}{500}\sum_{j=((t-500)+1)}^{((t-500)+500)}\|\lambda\|_{1,j}$ the mean of Y (t).
- Given $\text{Y}(t) = \left\{\|\lambda\|_{1,((t-500)+1)}, \cdots, \|\lambda\|_{1,((t-500)+500)}\right\}$, we defined its discrete Fourier transformation as $F_{k,t} = \sum_{j=((t-500)+1)}^{((t-500)+500)}\|\lambda\|_{1,j}\cdot e^{-2\pi ikj/500}$, where $k \in \{1, \cdots, 500\}$. The power spectrum is $PS_{k,t} = \left|F_{k,t}\right|^2$ for $k \in \{1, \cdots, 500\}$. Then, the MPS value at the date t, denoted as mps_t is the mean of all $PS_{k,t}$ for $k \in \left\{2, \cdots, {}^{1}\!/\!_{8} \times 500\right\}$.

As a result, we obtained three sets of time series based on AC1, VAR, and MPS values, which can be denoted as $AC1 = \{ac1_t | t = 549, \cdots, 1000\}$, $VAR = \{\text{var}_t | t = 549, \cdots, 1000\}$, and $MPS = \left\{mps_t | t = 549, \cdots, 1000\right\}$ accordingly.

Correlation Tests

In the study, Kendall's tau correlation was used to determine a rising trend in the indicator's time series (AC1, VAR, or MPS). Nonetheless, Pearson's and Spearman's rho correlations were included in addition to Kendall's tau correlation as alternative measures to detect the trend. In brief, Kendall's tau, Pearson's, and Spearman's rho correlations computed the strength of concordance dependency, the range of linear relationship, and the degree of the association, respectively.

Daily sliding window of length 250 was used to attain sequences containing 125 indicator values. In addition, the length of 250 was chosen as half data of the previous sliding window with length of 500. This length was considered sufficient to capture rising trend in the indicators as shown in Guttal et al. [21] and Ismail et al. [26]. Therefore, three tests based on Kendall's tau, Pearson's, and Spearman's rho correlations from each sequence containing 250 indicator values were computed. Considering AC1 as an example, each sequence containing 250 AC1 values at the date t can be denoted as $AC1(t) = \{ac1_t | t = ((t-250)+1), \cdots, ((t-250)+250)\}$, for $t = 798, \cdots, 1000$. For each $t \in \{798, \cdots, 1000\}$, Kendall's tau, Pearson's, and Spearman's rho correlations from $AC1(t)$ are computed as follow:

- $\tau_{AC1,t} = {}^{(C-D)}\!/\!_{Z}$, where C is the number of concordant pairs between $AC1(t)$ and $\{((t-250)+1), \cdots, ((t-250)+250)\}$, D is the number of discordant pairs between [[Mathtype-mtef1-eqn-136.mtf]] and $\{((t-250)+1), \cdots, ((t-250)+250)\}$. $Z = (250 \times (250-1))/2$ is the total number of different possible pair combinations.
- $r_{AC1,t} = \left(250\sum_{l=((t-250)+1)}^{((t-250)+250)} ac1_l \cdot l - \left(\sum_{l=((t-250)+1)}^{((t-250)+250)} ac1_l\right)\left(\sum_{l=((t-250)+1)}^{((t-250)+250)} l\right)\right) \Bigg/ \left(\sqrt{250\left(\sum_{l=((t-250)+1)}^{((t-250)+250)}(ac1_l)^2\right) - \left(\sum_{l=((t-250)+1)}^{((t-250)+250)} ac1_l\right)^2}\sqrt{250\left(\sum_{l=((t-250)+1)}^{((t-250)+250)}(l)^2\right) - \left(\sum_{l=((t-250)+1)}^{((t-250)+250)} l\right)^2}\right)$
- $p_{AC1,t} = 1 - \frac{6\sum_i d_i^2}{250(250^2-1)}$, where d_i is the difference between the ranks of the corresponding variables of $AC1(t)$, and $\{((t-250)+1), \cdots, ((t-250)+250)\}$.

As a result, three sets of time series based on Kendall's tau, Pearson's, and Spearman's rho correlation values computed from AC1 were obtained, which can be denoted as $\tau_{AC1} = \left\{\tau_{AC1,t} | t = 798, \cdots, 1000\right\}$, $r_{AC1} = \left\{r_{AC1,t} | t = 798, \cdots, 1000\right\}$, and $p_{AC1} = \left\{p_{AC1,t} | t = 798, \cdots, 1000\right\}$., respectively. For, VAR and MPS, the same three sets of time series based on these three correlation values also can be obtained using the above formula.

Note that if any correlation (Kendall's tau, Pearson's, or Spearman's rho) provides a positive real value at the date t, we conclude that a rising trend happened in the 250 indicator values from the date $(t-250+1)$ to t. In critical transition theory, it was expected that these indicators' time series: AC1, VAR, and MPS are increasing before a financial crisis. In addition, the rising trend in the indicators will give rise to a positive real number of the correlation ahead of a financial crisis.

Significant and Skewness Tests

Furthermore, we performed a significance test in this study to examine whether the rising trend in the last 125 indicator values at the date from 750 to 999, which was summarized by the correlation value at the date $t = 999$ (1 day before the financial crisis), was statistically significant at level 5%. To conduct the test, we computed the p-value for each of the correlation values at the date $t = 999$. The p-value obtained is a doubled probability of getting the statistical value or a value with even greater evidence against H_0.

Let us take AC1 as an example; here, H_0 represents no monotonic trend (either rising or declining) that happens in the time series $AC1(999)$. A significance test at level 5% translates to a requirement p-value less than 0.005 to interpret any monotonic trend (rising or decline) in the time series $AC1(999)$ as significant or otherwise. The same significant test was also applied to VAR and MPS time series, that is, $VAR(999)$ and $MPS(999)$, respectively. This significance test is vital to verify whether PH *via* L^1-normtime series associated with the used indicator (AC1, VAR, or MPS) can provide a reliable EWS at least 1 day before the observed financial crisis.

In addition, since the Pearson correlation is usually applied to time series whose probability distribution is symmetric, such as the normal or Gaussian probability distribution, it is better to consider other association correlation measures (Kendall's tau or Spearman's rho) when the probability distribution for the considered time series is asymmetric. Therefore, we also employed the skewness test to observe whether the probability distribution of the last 125 indicator values at the date from 750

to 999 is asymmetric. As a result, we also computed the skewness value for each of the Pearson correlation at the date $t = 999$.

If the skewness value equals zero, the observed series is perfectly symmetrical. But a skewness of exactly zero is quite unlikely for real-world data, so we interpreted the skewness value as follows:

- If skewness is between $-1/2$ and $1/2$, the distribution can be called approximately symmetric. For this case, the Pearson correlation is applicable to detect EWSs of impending financial crises.
- Otherwise, the distribution is called skewed, and the Pearson correlation is not considered.

Threshold

In this study, by using the test (Kendall's tau, Pearson's, or Spearman's rho correlation), if there exist two significant correlation values, which indicate the significant rising trend in the observed indicator (AC1, VAR, or MPS) 1 day before the two corresponding observed financial crises (Dotcom crash and Lehman Brothers bankruptcy) in the US, a threshold was determined by choosing the smaller value of those two values. Let us assume that T_1 and T_2 are these two thresholds for the US market, then $T = \min\{T_1, T_1\}$ is the threshold covering all available dates. The minimum value was chosen because it corresponded to the most extreme cases, which provided us with the longest period of significant rising trends in the correlation time series (AC1, VAR, and MPS) before the Dotcom crash and Lehman Brothers bankruptcy.

After computing the threshold, all available dates were covered, and the likelihood of an indicative period of significant rising trends, which lay above the threshold, was determined. The latter was made following the range of time at which the indicative trend could be observed. Finally, all the recorded events (periods with significant rising trends and breakpoints or without that signal) were classified either as EWSs, false alarms (FAs), false negatives (FNs), or true negatives (TNs) as described in the later section.

Classification

In our practice, we classified any recorded event (a period with significant rising trends and breakpoints or without that signal) as a EWS, an FA, an FN, or a TN and is as follows:

- If there is a nearest financial crisis that happened within a period of continuous significant rising trends with breakpoints, which lies above the threshold, we consider the event in this period as an EWS.
- If there is no nearest financial crisis that happened within a period of continuous significant rising trends with breakpoints, which lies above the threshold, the event on this period is considered as an FA.
- If there is a financial crisis that happened within a period without the observed signal, which lies below the threshold, we consider the event in this period as an FN.
- If there is no nearest financial crisis within a period without the observed signal, which lies below the threshold, the event in this period is considered as a TN.

TABLE 1 | The classification matrix used in this study.

		Does there exist a nearest financial crisis within the duration?	
		Exist	**None**
Has the correlation duration exceeded the threshold?	Yes	EWS	FA
	No	FN	TN

Evaluation Measures

For evaluation measures, the classification matrix used in our study is as given in **Table 1**. To evaluate the method's performance, we used two classification scores (in percentage), namely probability of successful anticipation (PSA) and probability of erroneous anticipation (PEA). These two scores are defined as follows: let A, B, C, and D be the total number of EWS, FA, FN, and TN, respectively, then

$$PSA = ((A + D)/(A + B + C + D)\ ) \times 100\% \quad (7)$$

$$PEA = ((B + C)/(A + B + C + D)\ ) \times 100\% \quad (8)$$

To evaluate, a method that scores the highest PSA and the lowest PEA is considered the best EWS detection tool.

RESULT

Two pre-crisis datasets were analyzed, whereby each of those datasets had a length of 1,000 days before a financial crisis. In the first row of **Figure 5**, the two pre-crisis datasets before the Dotcom crash and Lehman Brothers bankruptcy are presented, respectively. For all indices in each pre-crisis dataset, we computed all its corresponding daily log-return time series and then combined all the computed daily log returns to build a high-dimensional time series. Furthermore, the daily sliding window of length 50 was applied to obtain PCDs. Then, we applied PH on each PCD to obtain a corresponding L^1-norm value. As a result, we acquired a L^1-norm time series. All obtained L^1-norm time series for every pre-crisis dataset is illustrated in the second row of **Figure 5**.

Figure 5 demonstrates that the L^1-norm time series exhibits strong growth toward a primary peak prior to the Dotcom crash and Lehman Brothers bankruptcy in the US market. These results are consistent with Gidea and Katz [25]and Ismail et al. [26]. As stock indices become increasingly volatile when moving closer to a financial crisis, more peaks appear in its corresponding persistent landscape. The latter gives growth in the L^1-norm values prior to the Dotcom crash and Lehman Brothers bankruptcy [26].

Furthermore, the daily sliding window with the length of 500 was applied to each L^1-norm time series as shown in **Figure 5**, and the corresponding time series of three CSD indicators, including AC1, VAR, and MPS at low frequencies were obtained. Furthermore, for each computed CSD indicator's time series, we applied the daily sliding window with a length of 250 and

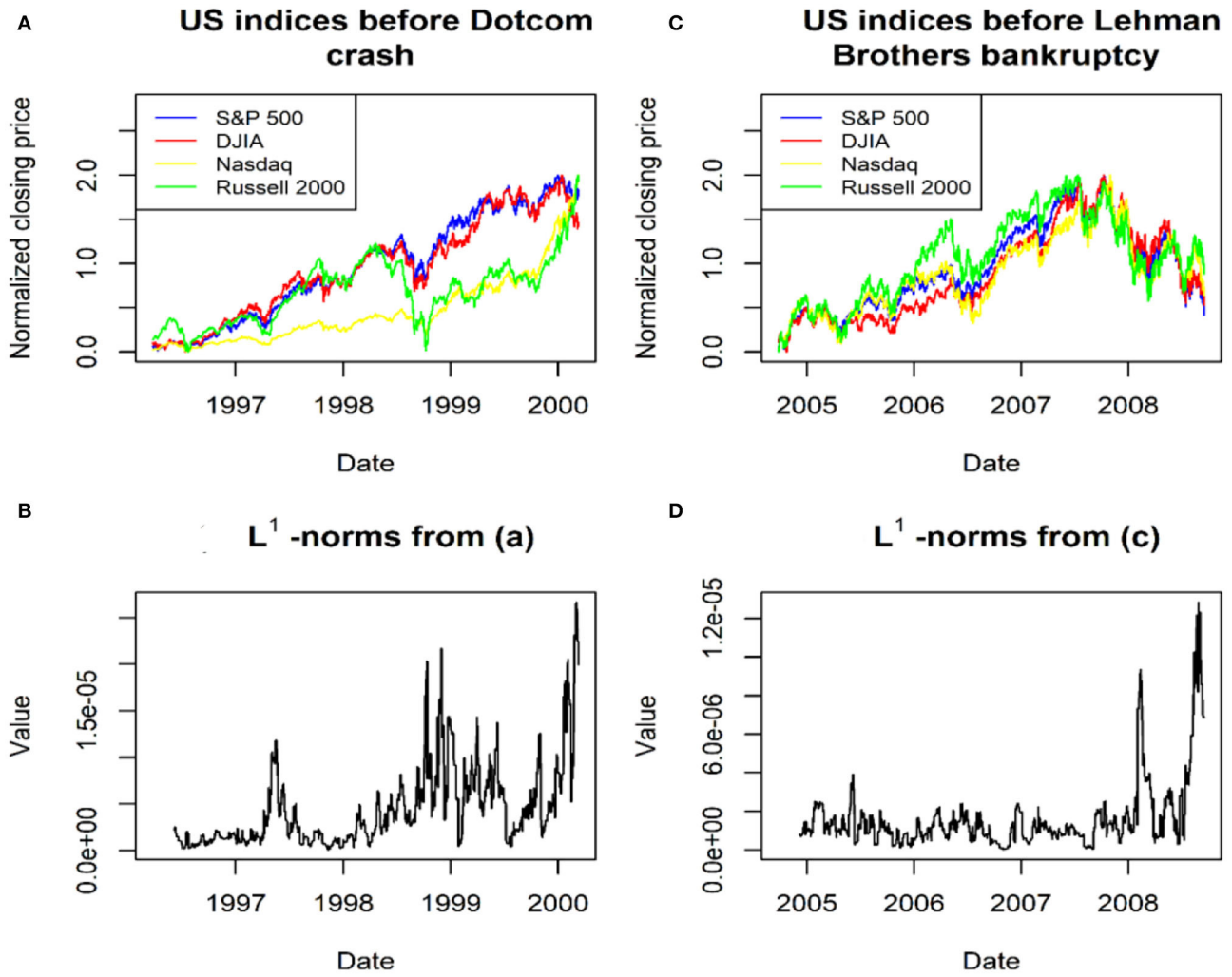


FIGURE 5 | (A,C) two pre-crisis datasets before the Dotcom crash and Lehman Brothers bankruptcy and **(B,D)** the two L^1-norm time series for these two datasets, respectively.

computed the time series of three tests: Kendall's tau, Pearson's, and Spearman's rho correlations. In addition, we also employed a significant test to compute p-values for each correlation test 1 day before the Dotcom crash and Lehman Brothers bankruptcy. **Figure 6** shows the time series of these CSD indicators (AC1, VAR, and MPS) and presented the obtained value for Kendall's tau, Pearson's, and Spearman's rho correlations before the financial crises. Moreover, in **Figure 6**, corresponding p-values for Kendall's tau, Pearson's, and Spearman's rho correlations and skewness measure (for the Pearson correlation only) are also mentioned. Kendall's tau correlation, Pearson's correlation, Spearman's rho correlation, p-values for Kendall's tau correlation, p-values for Pearson's correlation, p-values for Spearman's rho correlation, and skewness values in **Figure 6** are symbolized with τ, r, p, p_τ, p_r, p_p, and s, respectively.

Based on the results presented in **Figure 6**, for any CSD indicator of the L^1-normtime series, which consistently exhibits a significant rising trend (indicated by the correlation values, the corresponding p-values, and skewness) before the observed financial crises, we considered such method as a EWS detection tool. It is clearly shown that only the MPS time series of the L^1-norm time series fulfill this condition, and its significant rising trend can be indicated by Kendall tau, Pearson, and Spearman rho correlation. These because Kendall tau and Spearman rho provide positive values and their corresponding p-values also less than 0.05. For Pearson, this test obtains a positive value, its corresponding p-values is less than 0.05, and its skewness values lie in between $-1/2$ and $1/2$ indicating that the data has an approximately symmetrical distribution. For the others, they did not achieve the above-mentioned characteristics, therefore not considered as potential EWS detection tools.

In **Figure 6**, the highlighted values from different correlation tests (Kendall's tau, Pearson's, and Spearman's rho) are the chosen threshold to predict the whole available date. These values are 0.88, 0.94, and 0.96 for Kendall's tau, Pearson's, and Spearman's rho correlations, respectively. Therefore, we have

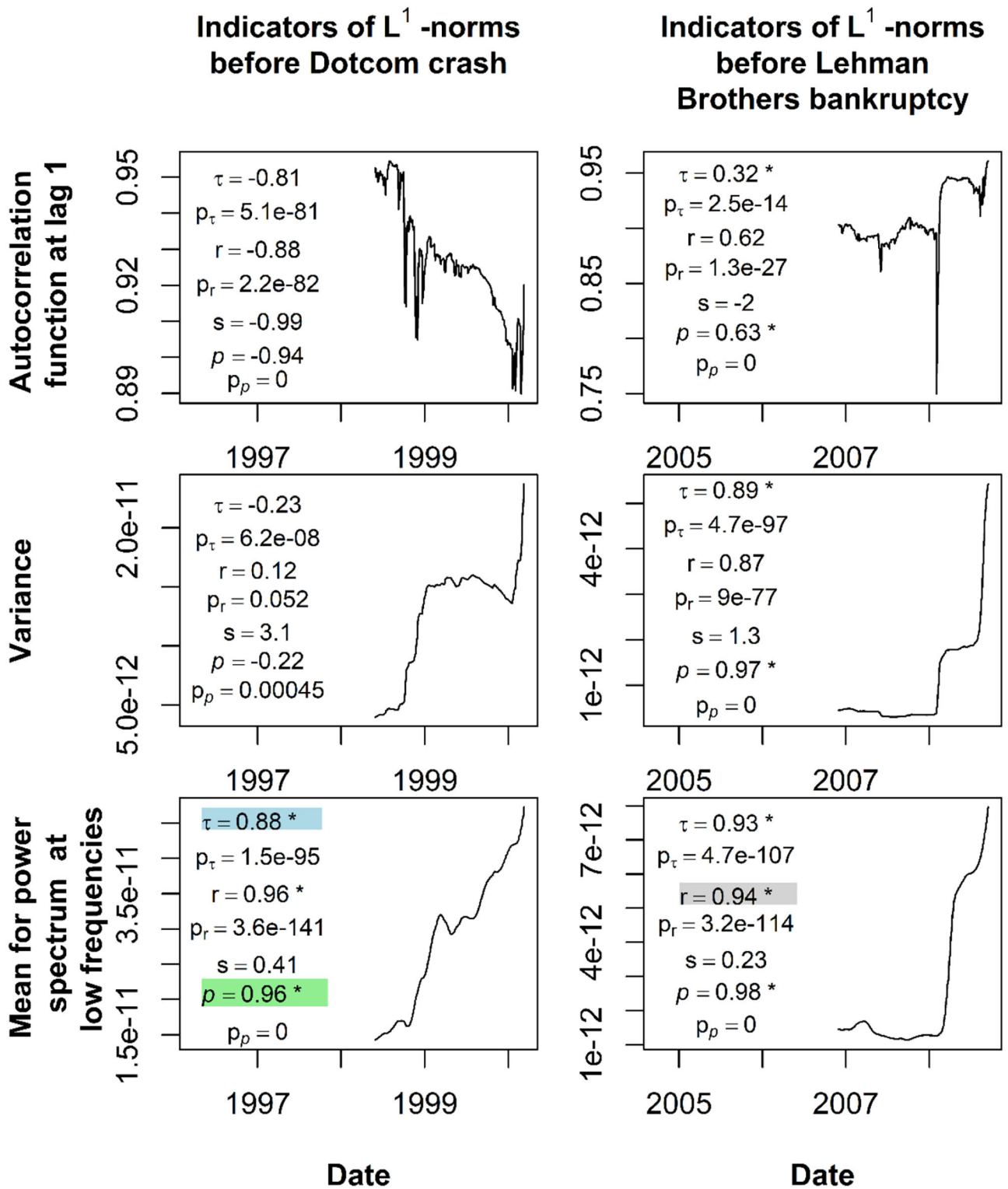


FIGURE 6 | Time series of the CSD indicators (AC1, VAR, and MPS), the obtained value for Kendall's tau, Pearson's, and Spearman's rho correlations, and their corresponding *p*-values 1 day before the financial crises.

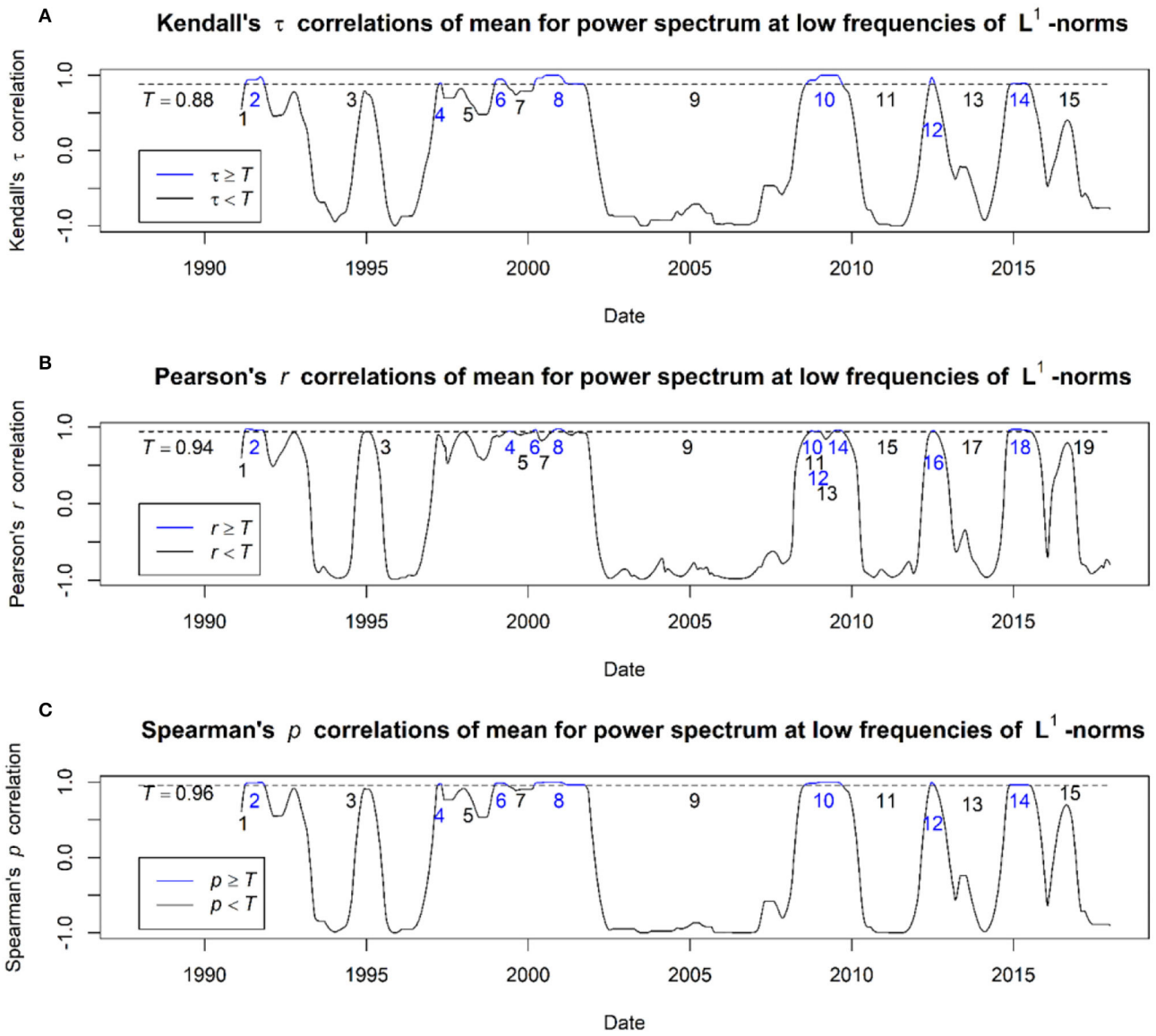


FIGURE 7 | Results obtained from MPS of the L^1-norm time series using Kendall's tau, Pearson's, and Spearman's rho correlations. **(A)** Kendall's tau correlations of mean for power spectrum at low frequencies of L^1-norms; **(B)** Pearson's rho correlations of mean for power spectrum at low frequencies of L^1-norms; **(C)** Spearman's rho correlations of mean for power spectrum at low frequencies of L^1-norms.

three predictions performed using three different correlations, and each correlation is obtained from the MPS time series of the L^1-normtime series. For each prediction, all individual events (periods of continuous significant rising trends or otherwise) are recorded, and these events are then classified whether they are EWSs, FAs, FNs, or TNs.

Results obtained from MPS of the L^1-normtime series using Kendall's tau, Pearson's, and Spearman's rho correlations, respectively, are illustrated in **Figure 7**. All periods of continuous significant rising trends in **Figure 7** are indicated by blue numbers and lines, which lie above the threshold. In contrast, correlation values that lie below the threshold are indicated with black numbers and lines. Details on these recorded events and their corresponding classification as shown in **Figure 7** for Kendall's tau, Pearson's, and Spearman's rho correlations are provided in tables of **Supplementary Material File**.

By using MPS of the L^1-norm time series, our results reported that Spearman's rho correlation test obtained the total number of EWS, FA, FN, TN, PSA, and PEA of 4, 7, 7, 8, 46.15, and 53.85%, respectively. Furthermore, Kendall's tau correlation test provided the total number of EWS, FA, FN, TN, PSA, and PEA of 3, 8, 7, 8, 42.31, and 57.69%, respectively. In addition, the Pearson correlation test achieved the total number of EWS, FA, FN, TN, PSA, and PEA of 2, 9, 9, 10, 40, and 60%, respectively.

TABLE 2 | A summary for all obtained results.

Evaluation measure	Kendall's tau correlation test	Pearson's correlation test	Spearman's rho correlation test
EWS	3	2	**4**
FA	8	9	**7**
FN	**7**	9	**7**
TN	**8**	10	**8**
PSA (%)	42.31	40	**46.15**
PEA (%)	57.69	60	**53.85**

All these results obtained are summarized in **Table 2**, where all the bolded values represent the obtained highest scores for the corresponding evaluation measure.

From **Table 2**, based on PSA and PEA, Spearman's rho and Kendall's tau correlations are showed to obtain a better result than the Pearson correlation. This shows that non-parametric rank correlation, which computes statistical associations based on the ranks of the data like Spearman's rho and Kendall's tau correlations is better than the Pearson correlation, which measures the degree of the linear relationship between related variables. However, Spearman's rho correlation, which measures the degree of association between two variables is reported and has a better result as compared to Kendall's tau correlation, which measures the strength of dependence between two variables. In addition, our result in **Table 2** clearly shows that Spearman's rho correlation outperforms other correlation tests (Kendall's tau and Pearson's correlations) with the highest score of every evaluation measure. Therefore, by using MPS of the L^1-norm time series, our results concluded that Spearman's rho correlation can detect EWSs better than Kendall's tau and Pearson's correlations.

CONCLUSION

In this study, PH and CSD were proposed to detect EWSs of major financial crashes in the US market. Preliminarily, two financial crises: Dotcom Crash and Lehman Brothers Bankruptcy that happened in the US market were examined. By using PH, L^1-norm time series was obtained for each financial crisis and used to compute the CSD indicators: AC1, VAR, and MPS.

By using three different correlation tests, Kendall's tau, Pearson's, and Spearman's rho, the rising trend in these indicators is observed prior to the financial crises. Furthermore, this study applied significance and skewness tests to determine whether the rising trends in the indicators (AC1, VAR, or MPS) are statistically significant. This test aims to conclude that the rising trend does not happen by chance. Subsequently, a threshold is determined to predict the whole date, and then the classification performance of our method is evaluated by using PSA and PEA.

Our result shows that the L^1-norm time series exhibits a strong growth before Dotcom Crash and Lehman Brothers Bankruptcy. This portrays that the L^1-norm time series has the potential to be a representative to detect EWSs of major financial crashes in the US market. Moreover, MPS from the L^1-norm time series is significantly rising before these two financial crises. It has also been demonstrated that all correlation tests, Kendall's tau, Pearson's, and Spearman's rho, can indicate the observed significant rising trend.

Overall, based on PSA and PEA, our results revealed that Spearman's rho correlation predicts the US market better than Kendall's tau and Pearson's correlations. Therefore, this study demonstrates that PH *via* its L^1-norm time series with MPS and Spearman's rho correlation offers a new potential EWS detection tool for financial crises in the US market. For future studies, we plan to examine and refine this method of PH to provide more reliable EWSs for upcoming financial crises.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://finance.yahoo.com

AUTHOR CONTRIBUTIONS

MSI: conceptualization, methodology, software, validation, writing–original draft, and writing–review and editing. MM: conceptualization, supervision, validation, writing–review and editing, and funding acquisition. MI: supervision, data curation, validation, writing–review and editing, funding acquisition. FA: supervision, validation, writing–review and editing. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We would like to express our gratitude to the Universiti Kebangsaan Malaysia and the Ministry of Higher Education Malaysia for their financial support *via* the provision of two grants with numbers: FRGS/1/2019/STG06/UKM/01/3 and GUP-2020-032.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fams.2022.940133/full#supplementary-material

REFERENCES

1. D.S. Bates. Jumps and stochastic volatility: exchange rate processes implicit in deutsche mark options. *Rev Financ Stud*. (2015) 9:69–107. doi: 10.1093/rfs/9.1.69
2. Gopalakrishnan EA, Sharma Y, John T, Dutta PS, Sujith RI. Early warning signals for critical transitions in a thermoacoustic system. *Sci Rep.* (2016) 6:35310. doi: 10.1038/srep35310
3. Gidea M, Goldsmith D, Katz Y, Roldan P, Shmalo Y. Topological recognition of critical transitions in time series of cryptocurrencies.

Physica A Stat Mech Appl. (2020) 548:123843. doi: 10.1016/j.physa.2019.123843
4. Virtanen T, Tölö E, Virén M, Taipalus K. Can bubble theory foresee banking crises? *J Financial Stab.* (2018) 36:66–81. doi: 10.1016/j.jfs.2018.02.008
5. Sornette D, Cauwels P. Financial bubbles: mechanisms and diagnostics. *Swiss Finance Institute Research Paper.* (2014). doi: 10.2139/ssrn.2423790
6. H.J. Edison. Do indicators of financial crises work? An evaluation of an early warning system. *Int J Finance Econ.* (2003) 8:11–53. doi: 10.1002/ijfe.197
7. Hubrich K, Tetlow RJ. Financial stress and economic dynamics: The transmission of crises. *J Monet Econ.* (2015) 70:100–15. doi: 10.1016/j.jmoneco.2014.09.005
8. Quax R, Kandhai D, Sloot P. Information dissipation as an early-warning signal for the Lehman Brothers collapse in financial time series. *Sci Rep.* (2013) 3:1–7. doi: 10.1038/srep01898
9. Gatfaoui H, Peretti PDe. Flickering in information spreading precedes critical transitions in financial markets. *Sci Rep.* (2019) 9:1–11. doi: 10.1038/s41598-019-42223-9
10. Squartini T, Van Lelyveld I, Garlaschelli D. Early-warning signals of topological collapse in interbank networks. *Sci Rep.* (2013) 3:1–9. doi: 10.1038/srep03357
11. Saracco F, Clemente RDi, Gabrielli A, Squartini T. Detecting early signs of the 2007–2008 crisis in the world trade. *Sci Rep.* (2016) 6:1–11. doi: 10.1038/srep30286
12. Almog A, Shmueli E. Structural entropy: monitoring correlation-based networks over time with application to financial markets. *Sci Rep.* (2019) 9:1–13. doi: 10.1038/s41598-019-47210-8
13. Flood MD, Lemieux VL, Varga M, Wong BW. The application of visual analytics to financial stability monitoring. *J Financial Stab.* (2016) 27:180–97. doi: 10.1016/j.jfs.2016.01.006
14. Battiston S, Farmer JD, Flache A, Garlaschelli D, Haldane AG, Heesterbeek H, et al. Complexity theory and financial regulation. *Science.* (2016) 351:818–9. doi: 10.1126/science.aad0299
15. D. Sornette, *Why Stock Markets Crash: Critical Events in Complex Financial Systems*. Princeton: University Press (2017). doi: 10.23943/princeton/9780691175959.001.0001
16. Van Nes EH, Scheffer M. Slow recovery from perturbations as a generic indicator of a nearby catastrophic shift. *Am Nat.* (2007) 169:738–47. doi: 10.1086/516845
17. Dakos V, Carpenter SR, Brock WA, Ellison AM, Guttal V, Ives AR, et al. Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data. *PLoS ONE.* (2012) 7:e41010. doi: 10.1371/journal.pone.0041010
18. Veraart AJ, Faassen EJ, Dakos V, van Nes EH, Lürling M, Scheffer M. Recovery rates reflect distance to a tipping point in a living system. *Nature.* (2012) 481:357–9. doi: 10.1038/nature10723
19. Tan JPL, Cheong SSA. Critical slowing down associated with regime shifts in the US housing market. *Eur Phys J B.* (2014) 87:1–10. doi: 10.1140/epjb/e2014-41038-1
20. Tan J, Cheong SA. The regime shift associated with the 2004–2008 US housing market bubble. *PLoS ONE.* (2016) 11:e0162140. doi: 10.1371/journal.pone.0162140
21. Guttal V, Raghavendra S, Goel N, Hoarau Q. Lack of critical slowing down suggests that financial meltdowns are not critical transitions, yet rising variability could signal systemic risk. *PLoS ONE.* (2016) 11:e0144198. doi: 10.1371/journal.pone.0144198
22. Wen H, Ciamarra MP, Cheong SA. How one might miss early warning signals of critical transitions in time series data: A systematic study of two major currency pairs. *PLoS ONE.* (2018) 13:e0191439. doi: 10.1371/journal.pone.0191439
23. Diks C, Hommes C, Wang J. Critical slowing down as an early warning signal for financial crises? *Empir Econ.* (2019) 57:1201–28. doi: 10.1007/s00181-018-1527-3
24. Otter N, Porter MA, Tillmann U, Grindrod P, Harrington HA, A. roadmap for the computation of persistent homology. *EPJ Data Science.* (2017) 6:1–38. doi: 10.1140/epjds/s13688-017-0109-5
25. Gidea M, Katz Y. Topological data analysis of financial time series: Landscapes of crashes. *Phys A Stat Mech Appl.* (2018) 491:820–34. doi: 10.1016/j.physa.2017.09.028
26. Ismail MS, Noorani MSM, Ismail M, Razak FA, Alias MA. Early warning signals of financial crises using persistent homology. *Phys A Stat Mech Appl.* (2022) 586:126459. doi: 10.1016/j.physa.2021.126459
27. Aromi LL, Katz YA, Vives J. Topological features of multivariate distributions: Dependency on the covariance matrix. *Commun Nonlinear Sci Numer Simuls.* (2021) 103:105996. doi: 10.1016/j.cnsns.2021.105996
28. M. Gidea, Topological Data Analysis of Critical Transitions in Financial Networks. In: Shmueli E, Barzel B, Puzis R, (Eds.), 3rd *International Winter School and Conference on Network Science*. Cham: Springer International Publishing (2017). pp. 47–59. doi: 10.1007/978-3-319-55471-6_5
29. Guo H, Xia S, An Q, Zhang X, Sun W, Zhao X. Empirical study of financial crises based on topological data analysis. *Phys A Stat Mech Appl.* (2020) 558:124956. doi: 10.1016/j.physa.2020.124956
30. Guo H, Zhao X, Yu H, Zhang X. Analysis of global stock markets' connections with emphasis on the impact of COVID-19. *Phys A Stat Mech Appl.* (2021) 569:125774. doi: 10.1016/j.physa.2021.125774
31. Yen PT, Xia K, Cheong SA. Understanding changes in the topology and geometry of financial market correlations during a market crash. *Entropy.* (2021) 23:1211. doi: 10.3390/e23091211
32. Yen PT, Cheong SA. Using Topological Data Analysis (TDA) and persistent homology to analyze the stock markets in Singapore and Taiwan. *Front Physics.* (2021) 9:20. doi: 10.3389/fphy.2021.572216
33. Kim W, Kim Y-J, Lee G, Kook W. Investigation of flash crash *via* topological data analysis. *Topol Appl.* (2021) 301:107523. doi: 10.1016/j.topol.2020.107523
34. Nguyen NKK, Bui M. Detecting anomalies in the dynamics of a market index with topological data analysis. *Int J Syst Innovation.* (2021) 6:37–50.
35. Katz YA, Biem A. Time-resolved topological data analysis of market instabilities. *Phys A Stat Mech Appl.* (2021) 571:125816. doi: 10.1016/j.physa.2021.125816
36. Ismail MS, Hussain SI, Noorani MSM. Detecting early warning signals of major financial crashes in bitcoin using persistent homology. *IEEE Access.* (2020) 8:202042–57. doi: 10.1109/ACCESS.2020.3036370
37. Goel A, Pasricha P, Mehra A. Topological data analysis in investment decisions. *Expert Syst Appl.* (2020) 147:113222. doi: 10.1016/j.eswa.2020.113222
38. Baitinger E, Flegel S. The better turbulence index? Forecasting adverse financial markets regimes with persistent homology. *Financ Mark Portf Manag.* (2021) 35:277–308. doi: 10.1007/s11408-020-00377-x
39. Majumdar S, Laha AK. Clustering and classification of time series using topological data analysis with applications to finance. *Expert Syst Appl.* (2020) 162:113868. doi: 10.1016/j.eswa.2020.113868
40. Ismail MS, Md Noorani MS, Ismail M, Abdul Razak F, Alias MA. Predicting next day direction of stock price movement using machine learning methods with persistent homology: Evidence from Kuala Lumpur Stock Exchange. *Appl Soft Comput.* (2020) 93:106422. doi: 10.1016/j.asoc.2020.106422
41. E. Baitinger, and S. Flegel, New Concepts in Financial Forecasting: Network-Based Information, Topological Data Analysis and their Combination. (2021). doi: 10.2139/ssrn.3962148
42. V. Robins, Towards computing homology from finite approximations. *J Adv Stud Topol.* (1999) 24: 503–32. Available online at: http://topology.nipissingu.ca/tp/reprints/v24/tp24222.pdf
43. Edelsbrunner H, Letscher D, Zomorodian A. Topological persistence and simplification. In: *Proceedings 41st Annual Symposium on Foundations of Computer Science*. New York, NY: IEEE (2000). p. 454–63.
44. Zomorodian A, Carlsson G. Computing persistent homology. *Discrete Comput Geom.* (2005) 33:249–74. doi: 10.1007/s00454-004-1146-y
45. G. Carlsson. Topology and data. *Bull New Ser Am Math Soc.* (2009) 46:255–308. doi: 10.1090/S0273-0979-09-01249-X
46. Edelsbrunner H, Harer JL, *Computational Topology: An Introduction*. American Mathematical Society (2022).
47. Bubenik P, Dłotko P, A. persistence landscapes toolbox for topological statistics. *J Symb Comput.* (2017) 78:91–114. doi: 10.1016/j.jsc.2016.03.009

48. Bubenik P. Statistical topological data analysis using persistence landscapes. *J Mach Learn Res.* (2015) 16:77–102. doi: 10.48550/arXiv.1207.6437

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor MM declared a past collaboration with the authors.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Applied Mathematics and Statistics

TYPE Original Research
PUBLISHED 24 August 2022
DOI 10.3389/fams.2022.958081

OPEN ACCESS

EDITED BY
Mohd Hafiz Mohd,
Universiti Sains Malaysia (USM), Malaysia

REVIEWED BY
Amirah Azmi,
Universiti Sains Malaysia (USM), Malaysia
Syaza Latif,
MARA University of Technology, Malaysia

*CORRESPONDENCE
Nursanti Anggriani
nursanti.anggriani@unpad.ac.id

SPECIALTY SECTION
This article was submitted to Mathematics of Computation and Data Science, a section of the journal Frontiers in Applied Mathematics and Statistics

RECEIVED 31 May 2022
ACCEPTED 05 July 2022
PUBLISHED 24 August 2022

CITATION
Inayaturohmat F, Anggriani N and Supriatna AK (2022) A mathematical model of tuberculosis and COVID-19 coinfection with the effect of isolation and treatment.
Front. Appl. Math. Stat. 8:958081.
doi: 10.3389/fams.2022.958081

A mathematical model of tuberculosis and COVID-19 coinfection with the effect of isolation and treatment

Fatuh Inayaturohmat[1], Nursanti Anggriani[2]* and Asep K. Supriatna[2]

[1]Master of Mathematics Study Program, Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Sumedang, Indonesia, [2]Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Sumedang, Indonesia

In this research, we developed a coinfection model of tuberculosis and COVID-19 with the effect of isolation and treatment. We obtained two equilibria, namely, disease-free equilibrium and endemic equilibrium. Disease-free equilibrium is a state in which no infection of tuberculosis and COVID-19 occurs. Endemic equilibrium is a state in which there occurs not only the infection of tuberculosis and COVID-19 but also the coinfection of tuberculosis and COVID-19. We assumed that the parameters follow the uniform distribution, and then, we took 1,000 samples of each parameter using Latin hypercube sampling (LHS). Next, the samples were sorted by ranking. Finally, we used the partial rank correlation coefficient (PRCC) to find the correlation between the parameters with compartments. We analyzed the PRCC for three compartments, namely, individuals infected with COVID-19, individuals infected with tuberculosis, and individuals coinfected with COVID-19 and tuberculosis. The most sensitive parameters are the recovery rate and the infection rate of each COVID-19 and tuberculosis. We performed the optimal control in the form of prevention for COVID-19 and tuberculosis. The numerical simulation shows that these controls effectively reduce the infected population. We also concluded that the effect of isolation has an immediate impact on reducing the number of COVID-19 infections, while the effect of treatment has an impact that tends to take a longer time.

KEYWORDS

mathematical model, tuberculosis, COVID-19, coinfection, isolation, treatment, PRCC, optimal control

Introduction

A healthy and prosperous life is one of the things agreed upon by countries in the world, including Indonesia. Efforts to promote healthy living and support welfare for all ages are listed in the third point of the Sustainable Development Goals (SDGs). These goals include several other things, including reducing the ratio of deaths from the disease to <70 per 100,000 live births, developing treatments, and defining the AIDS, tuberculosis, and malaria epidemics [1].

Tuberculosis (TB) is one of the most deadly infectious diseases globally. Even individuals under treatment for tuberculosis are still infectious for some time

(different for each individual), especially when the bacteria are still active [2]. Globally, more than 4,100 people die daily from TB, and almost 28,000 people are infected with this preventable and curable disease. Global efforts to fight TB have saved about 66 million lives since 2000. However, the COVID-19 pandemic has reversed years of progress made in the fight to end TB. For the first time in a decade, TB deaths increased in 2020. Globally, TB incidence decreased by about 2% annually in 2015–2020 with a cumulative 11%. Due to the COVID-19 pandemic, nearly 9.9 million people were infected, and nearly 1.5 million people died from TB in 2020 [3]. In 2021, there were 443,235 tuberculosis cases in Indonesia. Until 2 June 2022, 155,822 cases of tuberculosis were confirmed [4].

COVID-19 is a disease that spreads very quickly, causing a pandemic and becoming a global health problem. The pandemic has had a major impact on other sectors, especially the socioeconomic sector [5]. Based on the Worldometer, as of 30 April 2022, there were 512,466,045 positive cases of COVID-19 in the world, with 6,257,512 deaths. Indonesia is ranked 18th in the world and 7th in Asia with 6,046,467 total positive cases of COVID-19 [6].

Based on these data, efforts are needed to reduce transmission and prevent the spread of tuberculosis and COVID-19. Mathematics has an essential role in modeling the epidemic phenomenon of the disease, one of which is by using a deterministic model of the spread of the disease. The study of the COVID-19 model with waning immunity has undergone various developments. Models with symptomatic and asymptomatic infected populations show that, in determining the beginning time of the massive vaccination strategy, the vaccine efficacy level is a deciding factor, whereas the natural immunity period is for the latter [7]. Vaccination and treatment are very effective in suppressing the spread of COVID-19 [8, 9]. In addition, isolation and vaccination are effective strategies for reducing the infection [10]. The effort to control the multiple COVID-19 strains is the demand for more restrictive measures [11]. The SEIR model with vaccination and pre-symptomatic populations shows that the waning of vaccine-induced immunity significantly impacts the disease spreading [12]. Contact tracing, national-scale social restriction, wearing face masks, and case detection and treatment are the most effective scenarios to control the spreading of COVID-19 in Indonesia [13]. The optimal control of the SEIR model in Nigeria can be effective if the optimal control implemented is capable of sustaining the basic reproduction number to below unity [14]. COVID-19 reinfection scenario with the SIR model in Malaysia shows that transmission dynamics could emerge due to the waning of immunity, even when the reinfection force is relatively low [15]. The model of COVID-19 with infection through goods contaminated with SARS-CoV-2 suggests that the public should be aware [16]. The COVID-19 model with a healthcare population shows the importance of the protection of healthcare workers [17]. In the COVID-19 model with an educated population and contaminated surface objects, in the absence of a vaccine, countries need to detect and isolate at least 30% of the asymptomatic individuals while treating at least 50% of symptomatic individuals to control COVID-19 [18]. The SEIR model with isolation and hospitalization shows that the enhanced government, individual-level interventions, and the intensity of media coverage could substantially suppress the transmission of COVID-19 cases in Ghana [19]. A researcher has started to develop the coinfection model for tuberculosis in the Middle East [20].

Researchers also constructed a COVID-19 model with the effect of comorbidity on the population [21] and also the effect of home care with non-pharmaceutical interventions and sick vaccinations [22]. It was reported that the authors [23] developed a model with a partial comorbid population, while the other authors [24] examined a comorbid and isolated COVID-19 model. Some researchers did their research on the model of tuberculosis and HIV coinfection with optimal control and sensitivity analysis [25], the model of tuberculosis and HIV coinfection with the effect of treatment [26], the model of tuberculosis and HIV coinfection with treatment interventions [27], the effect of HIV immunity on treatment control [28], the model of tuberculosis and HIV coinfection with the effect of saturated care [29], the effect of reinfection on the model of tuberculosis and HIV coinfection [30], the relationship between tuberculosis and HIV and diabetes using a fractional differential equation model [31], the drug resistance from viruses that cause tuberculosis due to non-treatment [32], and the development of prevention strategies of spread in an AIDS-influenced model of tuberculosis [33].

The total number of infections and deaths by tuberculosis has increased due to COVID-19 [3]. Tuberculosis and COVID-19 can infect a human at the same time. Therefore, few researchers developed models of coinfection from tuberculosis and COVID-19. The model forecasts the dynamics of COVID-19 and tuberculosis in Delhi, India [34]. In the differential fractional equation for tuberculosis and COVID-19 coinfection using the Atangana-Baleanu derivative, the result is that reducing COVID-19 infection by tuberculosis-infected individuals can reduce the spread of infection and coinfection in the population [35]. By taking into account the reported and unreported cases in the model of tuberculosis and COVID-19 coinfection, it was determined that the best control used to reduce the spread of tuberculosis and COVID-19 in the population is to focus on preventing COVID-19 at a minimum cost [36]. The result of this study is that reducing infection contacts can reduce the spread of tuberculosis and COVID-19 coinfection [37].

Based on the description above, the authors are interested in developing a mathematical model on the coinfection of tuberculosis and COVID-19 with the effect of isolation and treatment. Isolation is an effective strategy to reduce the spreading of COVID-19 besides vaccination because there is no specific cure yet for COVID-19 [10]. On the other hand, treatment for individuals infected with tuberculosis is

an effective strategy to reduce the spreading of tuberculosis because it can reduce the risk of transmitting tuberculosis from the infected individuals [26–28]. Dynamic analysis is used to study and analyze the behavior and dynamics of the spread of tuberculosis and COVID-19. The numerical simulations are used to support the results obtained.

Materials and methods

We developed a mathematical model by dividing the human population into seven compartments: susceptible individuals (S), individuals infected with COVID-19 (I_C), individuals infected with tuberculosis (I_T), individuals coinfected with COVID-19 and tuberculosis (I_{TC}), individuals isolated with COVID-19 infection (Q), individuals under tuberculosis treatment (T), and recovered individuals (R).

The model has the following assumptions:

i. Individuals infected with tuberculosis can get infected with COVID-19 and vice versa.
ii. Individuals coinfected with tuberculosis and COVID-19 can transmit either tuberculosis or COVID-19.
iii. Coinfected individuals can recover from either COVID-19 or tuberculosis and the mixed infection at the same time.
iv. There are two different infection rates, namely, the infection rate of tuberculosis (γ_T) and the infection rate of COVID-19 (γ_C). Infection rates for singly infected and coinfected individuals are assumed to be the same.
v. Isolated individuals cannot transmit COVID-19, but individuals under tuberculosis treatment can transmit tuberculosis.
vi. The recovery rates for individuals infected with tuberculosis, infected with COVID-19, coinfected with tuberculosis and COVID-19, isolated because of the COVID-19 infection, and under tuberculosis treatment are different.
vii. Coinfected individuals have a higher death rate than singly infected individuals.

Human populations are recruited into susceptible individuals with the recruitment rate A. The authors assumed that natural death is the same for all compartments with the natural death rate μ. Susceptible individuals decrease due to the tuberculosis infection following effective contact with either singly or coinfected individuals at the rate:

$$\beta_T = \frac{\gamma_T (I_T + I_{TC} + T)}{N}$$

Susceptible individuals also decrease due to the COVID-19 infection after effective contact with either singly or coinfected individuals at the rate:

$$\beta_C = \frac{\gamma_C (I_C + I_{TC})}{N}$$

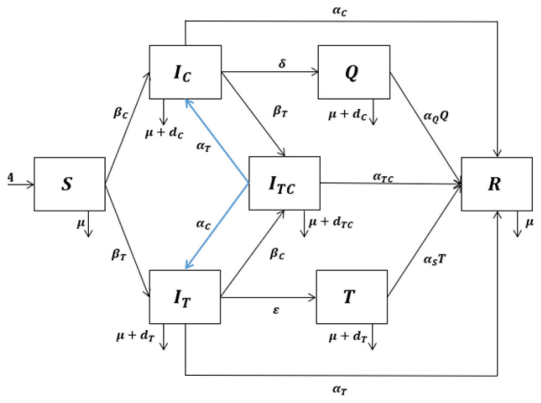


FIGURE 1
Transmission diagram.

TABLE 1 Description of the compartments.

Compartment	Description
S	Susceptible individuals
I_C	Individuals infected with COVID-19
I_T	Individuals infected with tuberculosis
I_{TC}	Individuals coinfected with COVID-19 and tuberculosis
Q	Individuals isolated with COVID-19 infection
T	Individuals under tuberculosis treatment
R	Recovered individuals

The transmission diagram of the model can be seen in Figure 1. The system of the differential equation for the tuberculosis and COVID-19 coinfection model with the effect of isolation and treatment based on the assumptions and the above-given description is as follows:

$$\begin{aligned}
\frac{dS}{dt} &= A - (\beta_C + \beta_T + \mu)\ S(t) \\
\frac{dI_C}{dt} &= \beta_C S(t) + \alpha_T I_{TC}(t) - \left(\beta_T + \delta + \mu + \alpha_C + d_C\right)\ I_C(t) \\
\frac{dI_T}{dt} &= \beta_T S(t) + \alpha_C I_{TC}(t) - \left(\beta_C + \varepsilon + \mu + \alpha_T + d_T\right)\ I_T(t) \\
\frac{dI_{TC}}{dt} &= \beta_C I_T(t) + \beta_T I_C(t) \\
&\quad - \left(\alpha_{TC} + \mu + \alpha_T + \alpha_C + d_{TC}\right)\ I_{TC}(t) \\
\frac{dQ}{dt} &= \delta I_C(t) - \left(\alpha_Q + \mu + d_C\right) Q(t) \\
\frac{dT}{dt} &= \varepsilon I_T(t) - \left(\alpha_S + \mu + d_T\right) T(t) \\
\frac{dR}{dt} &= \alpha_C I_C(t) + \alpha_T I_T(t) + \alpha_Q Q(t) + \alpha_S T(t) \\
&\quad + \alpha_{TC} I_{TC}(t) - \mu R(t)
\end{aligned} \tag{1}$$

The description of the compartments and the values of parameters are provided in Tables 1, 2, respectively. Despite the

TABLE 2 Description and value of the parameters.

Parameter	Description	Value	Source
A	Recruitment rate	$\frac{10000}{59.365}$	[38]
μ	Natural death rate	$\frac{1}{59.365}$	[38]
α_T	Recovery rate of individuals infected with tuberculosis	$0.001 - 0.035$	Assumed
α_C	Recovery rate of individuals infected COVID-19	$0.001 - 0.03$	Assumed
α_{TC}	Recovery rate of individuals coinfected with tuberculosis and COVID-19	$0.001 - 0.01$	Assumed
α_Q	Recovery rate of COVID-19 isolated individuals	$0.001 - 0.3$	Assumed
α_S	Recovery rate of individuals under tuberculosis treatment	$0.001 - 0.4$	Assumed
δ	Rate of isolation	$0.001 - 0.5$	Assumed
ε	Rate of treatment	$0.001 - 0.2$	Assumed
γ_T	Infection rate of tuberculosis	2×10^{-6}	Assumed
γ_C	Infection rate of COVID-19	5.5×10^{-6}	Assumed
d_T	Death rate of tuberculosis	0.004	[39]
d_C	Death rate of COVID-19	0.0018	[40]
d_{TC}	Death rate of coinfected tuberculosis and COVID-19	$0.005 - 0.1$	Assumed
β_T	Force of infection of tuberculosis	–	–
β_C	Force of infection of COVID-19	–	–

unavailability of data to apply to this study, the various assumed parameters used in our model have demonstrated a behavior of the spread of tuberculosis. Some of these values are taken from previous studies as indicated in Table 2, and some of these values which are not available are assumed within plausible ranges.

Result and discussion

We analyzed the local stability of the equilibrium. Then, the basic reproduction number is generated from the disease-free equilibrium using the next-generation matrix method. Next, the global sensitivity analysis of the model is conducted to find the most sensitive parameter of the model and how significant the parameter change affects the model. In this research, we used the Latin hypercube sampling (LHS) to take samples and then the partial rank correlation coefficient (PRCC) to find the partial correlation between the parameters and the compartments. The most sensitive parameter will be controlled with the optimal control. Finally, we analyzed the dynamics of the population using the numerical simulation with and without control.

Dynamical analysis

Setting all the differential equations from the model (1) equaling zero and then solving for all compartments, we obtained two equilibria of the model, namely, disease-free equilibrium (E^0) and endemic equilibrium (E^*).

Disease-free equilibrium:

$$E^0 = \{S, I_C, I_T, I_{TC}, Q, T, R\} = \left\{\frac{A}{\mu}, 0, 0, 0, 0, 0, 0\right\} \tag{2}$$

Endemic equilibrium:

$$E^* = \left\{S^*, I_C^*, I_T^*, I_{TC}^*, Q^*, T^*, R^*\right\} \tag{3}$$

$$S^* = \frac{A}{\beta_C^* + \beta_T^* + \mu}$$

$$I_C^* = \frac{A\beta_C\left[(L2 + L3 + \alpha_C + \beta_C + \beta_T)\,\alpha_T + (\varepsilon + \mu + d_T)\,(\alpha_C + L3) + L3\beta_C\right]}{L4\left\{\left[\alpha_C^2 + (1 + L2 + L3 + \beta_C)\,L1 + \beta_T\,(L3 + \alpha_C)\right]\alpha_T + \left[\left((\varepsilon + \mu + d_T)\,(\alpha_C + L3) + L3\beta_C\right)(L1 + \beta_T)\right]\right\}}$$

$$I_T^* = \frac{A\,\beta_T\left[(L3 + \alpha_C)\,\beta_T + (L3 + L1 + \alpha_T + \beta_C)\,\alpha_C + (\delta + \mu + d_C)\,(L3 + \alpha_T)\right]}{L4\left\{[L2\alpha_C + (L2 + \beta_C)\,L3]\,\beta_T + L1\,[L2\alpha_C + (L2 + \beta_C)\,(L3 + \alpha_T)]\right\}}$$

$$I_{TC}^* = \frac{A\,\beta_C\beta_T\,(L1 + L2 + \beta_C + \beta_T)}{L4\left[(\beta_C L3 + L2\,(L3 + \alpha_C))\,(L1 + \beta_T) + L1\alpha_T(\beta_C + L2)\right]}$$

$$Q^* = \frac{\delta I_C^*}{(\alpha_Q + \mu + d_C)}$$

$$T^* = \frac{\varepsilon I_T^*}{(\alpha_S + \mu + d_T)}$$

$$R^* = \frac{\alpha_C I_C(t) + \alpha_T I_T(t) + \alpha_Q Q(t) + \alpha_S T(t) + \alpha_{TC} I_{TC}(t)}{\mu},$$

$$V = \begin{pmatrix} (\delta+\mu+\alpha_C+d_C) & 0 & -\alpha_T & 0 \\ 0 & (\varepsilon+\mu+\alpha_T+d_T) & -\alpha_C & 0 \\ 0 & 0 & (\alpha_{TC}+\mu+\alpha_T+\alpha_C+d_{TC}) & 0 \\ 0 & -\varepsilon & 0 & (\alpha_S+\mu+d_T) \end{pmatrix}$$

where

$$FV^{-1} = \begin{pmatrix} \frac{\gamma_C A}{\mu\theta_1} & 0 & \frac{\gamma_C A\alpha_T}{\mu\theta_1\theta_3} + \frac{\gamma_C A}{\mu\theta_3} & 0 \\ 0 & \frac{\gamma_T A}{\mu\theta_2} + \frac{\gamma_T A\varepsilon}{\mu\theta_2(\alpha_S+\mu+d_T)} & \frac{\gamma_T A}{\mu\theta_3}\left[1+\frac{\alpha_C}{\theta_2}\left(1+\frac{\varepsilon}{(\alpha_S+\mu+d_T)}\right)\right] & \frac{\gamma_T A}{\mu(\alpha_S+\mu+d_T)} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$L1 = (\delta+\alpha_C+\mu+d_C)$$
$$L2 = (\varepsilon+\alpha_T+\mu+d_T)$$
$$L3 = (\mu+d_{TC}+\alpha_{TC})$$
$$L4 = (\beta_C+\beta_T+\mu)$$

The basic reproduction number ($\mathfrak{R}_0$) is the expected number of secondary cases produced by a typical infected individual during its entire period of infectiousness [41, 42]. When $\mathfrak{R}_0 < 1$, the disease will disappear from the population. Otherwise, when $\mathfrak{R}_0 > 1$, the disease will remain or spread in the population.

Let $x = (I_C, I_T, I_{TC}, T)$, then the model can be written as follows:

$$\frac{dx}{dt} = \mathcal{F}(X) - \mathcal{V}(X),$$

where $\mathcal{V}(X) = \mathcal{V}^-(X) - \mathcal{V}^+(X)$. $\mathcal{F}(X)$ is the rate of appearance of new infections in compartments, $\mathcal{V}^+(X)$ is the rate of transfer of individuals into the compartment, and $\mathcal{V}^-(X)$ is the rate of transfer of individuals out of the compartment.

$$\mathcal{F}(X) = \begin{pmatrix} \beta_C S(t) \\ \beta_T S(t) \\ \beta_C I_T(t) + \beta_T I_C(t) \\ 0 \end{pmatrix}$$

$$\mathcal{V}(X) = \begin{pmatrix} -\alpha_T I_{TC}(t) + (\beta_T+\delta+\mu+\alpha_C+d_C)\, I_C(t) \\ -\alpha_C I_{TC}(t) + (\beta_C+\varepsilon+\mu+\alpha_T+d_T)\, I_T(t) \\ (\alpha_{TC}+\mu+\alpha_T+\alpha_C+d_{TC}) I_{TC}(t) \\ -\varepsilon I_T(t) + (\alpha_S+\mu+d_T)\, T(t) \end{pmatrix}$$

Substituting the disease-free equilibrium in the Jacobian matrix of $\mathcal{F}(X)$ and $\mathcal{V}(X)$, we obtained the following equation:

$$F = \begin{pmatrix} \frac{\gamma_C A}{\mu} & 0 & \frac{\gamma_C A}{\mu} & 0 \\ 0 & \frac{\gamma_T A}{\mu} & \frac{\gamma_T A}{\mu} & \frac{\gamma_T A}{\mu} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Hence, the next-generation matrix is as follows:

where

$$\theta_1 = \delta+\mu+\alpha_C+d_C$$
$$\theta_2 = \varepsilon+\mu+\alpha_T+d_T$$
$$\theta_3 = \alpha_{TC}+\mu+\alpha_T+\alpha_C+d_{TC}$$

The basic reproduction number is the spectral radius of FV^{-1}:

$$\mathfrak{R}_0 = \rho\left(FV^{-1}\right) = \max\left(\frac{\gamma_C A}{\mu(\delta+\mu+\alpha_C+d_C)}, \frac{\gamma_T A(\alpha_S+\varepsilon+\mu+d_T)}{\mu(\alpha_T+\varepsilon+\mu+dt)(\alpha_S+\mu+dt)}\right) \quad (4)$$

with

$$\mathfrak{R}_{0C} = \frac{\gamma_C A}{\mu(\delta+\mu+\alpha_C+d_C)} \quad (5)$$

$$\mathfrak{R}_{0T} = \frac{\gamma_T A(\alpha_S+\varepsilon+\mu+d_T)}{\mu(\alpha_T+\varepsilon+\mu+dt)(\alpha_S+\mu+dt)} \quad (6)$$

The basic reproduction number $\mathfrak{R}_{0C}$ and $\mathfrak{R}_{0T}$ represent the number of infection cases produced by one infectious COVID-19 case and by one infectious tuberculosis case, respectively. The basic reproduction number of the coinfection model (1) is given in Equation (4).

We analyzed the local stability of each equilibrium through its eigenvalues [43]. The eigenvalues are obtained from the Jacobian matrix, which has been substituted with each equilibrium. The Jacobian matrix of model (1) can be seen below:

$$J = \begin{pmatrix} -\beta_C - \beta_T - \mu & -\gamma_C S(t) & -\gamma_T S(t) & -(\gamma_C + \gamma_T) S(t) & 0 & -\gamma_T S(t) & 0 \\ \beta_C & \gamma_C S(t) - \beta_T - \theta_1 & -\gamma_T I_C(t) & -\gamma_T I_C(t) + \gamma_C S(t) + \alpha_T & 0 & -\gamma_T I_C(t) & 0 \\ \beta_T & -\gamma_C I_T(t) & \gamma_T S(t) - \beta_C - \theta_2 & -\gamma_C I_T(t) + \gamma_T S(t) + \alpha_C & 0 & \gamma_T S(t) & 0 \\ 0 & \gamma_C I_T(t) + \beta_T & \gamma_T I_C(t) + \beta_C & \gamma_C I_T(t) + \gamma_T I_C(t) - \theta_3 & 0 & \gamma_T I_C(t) & 0 \\ 0 & \delta & 0 & 0 & -\alpha_Q - \mu - d_C & 0 & 0 \\ 0 & 0 & \varepsilon & 0 & 0 & -\alpha_S - \mu - d_T & 0 \\ 0 & \alpha_C & \alpha_T & \alpha_{TC} & \alpha_Q & \alpha_S & -\mu \end{pmatrix} \quad (7)$$

Substituting the disease-free equilibrium in Equation (7), we obtained the following characteristic polynomial:

$$\mathcal{P}_0(\lambda) = (\lambda_1 + \mu)(\lambda_2 + \mu)(\lambda_3 + \theta_3)\left(\lambda_4 + \alpha_Q + d_C + \mu\right)\left(\lambda_5 - \frac{A\gamma_C}{\mu} + \theta_1\right)\left(k_2\lambda^2 + k_1\lambda + k_0\right) \quad (8)$$

From Equation (8), we obtained the following equation:

$$\lambda_1 = -\mu, \lambda_2 = -\mu, \lambda_3 = -\alpha_C - \alpha_T - \alpha_{TC} - d_{TC} - \mu, \lambda_4 = -\alpha_Q - d_C - \mu, \lambda_5 = \frac{A\gamma_C}{\mu} - \alpha_C - \delta - d_C - \mu.$$

It is clear that $\lambda_1, \lambda_2, \lambda_3, \lambda_4 < 0$. For λ_5, we obtained the following equation:

$$\frac{A\gamma_C}{\mu} - \alpha_C - \delta - \mu - d_C < 0$$
$$\frac{A\gamma_C}{\mu} < \alpha_C + \delta + \mu + d_C$$
$$\frac{A\gamma_C}{\mu(\alpha_C + \delta + \mu + d_C)} < 1$$
$$\Re_{0C} < 1$$

Then, λ_i with $i = 6, 7$ will be negative if $k_j > 0$ with $j = 0, 1, 2$. We numerically analyzed the coefficient values because the characteristic polynomial $\mathcal{P}_0(\lambda)$ coefficients are complex. The result of numerical analysis is for k_j with $j = 0, 1, 2$ (Appendix 1).

Since λ_i with $i = 1, 2, 3, \ldots, 7$ is negative, it can be concluded that the disease-free equilibrium is locally asymptotically stable when $\Re_0 < 1$.

By analyzing the stability of the endemic equilibrium by substituting the endemic equilibrium in Equation (7), we obtained the following characteristic polynomial:

$$\mathcal{P}^*(\lambda) = \left(m_7\lambda^7 + m_6\lambda^6 + m_5\lambda^5 + m_4\lambda^4 + m_3\lambda^3 + m_2\lambda^2 + m_1\lambda + m_0\right) = 0 \quad (9)$$

λ_i with $i = 1, 2, 3, \ldots, 7$ will be negative if $m_j > 0$ with $j = 0, 1, 2, \ldots, 7$. We numerically analyzed the coefficient values because the characteristic polynomial $\mathcal{P}^*(\lambda)$ coefficients are complex. The result of numerical analysis is m_j with $j = 0, 1, 2, \ldots, 7$ (Appendix 2).

Since λ_i with $i = 1, 2, 3, \ldots, 7$ is negative, it can be concluded that the endemic equilibrium is locally asymptotically stable when $\Re_0 > 1$.

Partial rank correlation coefficient

We performed a global sensitivity analysis of the model by using the combination of LHS and PRCC. The purpose of the global sensitivity analysis is to find the most sensitive parameter to the model. LHS divides the range of the sample and so we took samples from each partition evenly. PRCC gives a partial correlation of each parameter to the model. The correlation is from −1 to 1. If the correlation is negative, then it indicates that, when the parameter decreases, the compartment also decreases; otherwise, the correlation is positive [44]. We assumed that the parameters follow the uniform distribution $[U(0, 1)]$, and then, we took 1,000 samples using LHS. Next, the samples were sorted by ranking. Finally, we used PRCC to find the correlation between the parameters with compartments. In this research, we analyzed the PRCC for three compartments, namely, individuals infected with COVID-19 (I_C), individuals infected with tuberculosis (I_T), and individuals coinfected with COVID-19 and tuberculosis (I_{TC}). The result of the global sensitivity analysis for all parameters of the infected population is illustrated in Figures 2–4.

Figure 2 shows that the most sensitive parameters of the individuals infected with COVID-19 (I_C) are the recovery rate of individuals infected with COVID-19 (α_C), which has a negative correlation, and the infection rate of COVID-19 (γ_C), which has a positive correlation. Meanwhile, the isolation rate also has a high negative correlation with the individuals infected with COVID-19 (I_C). It means that the effort to isolate individuals infected with COVID-19 (I_C) can suppress the spreading of COVID-19.

Figure 3 shows that the most sensitive parameters of the individuals infected with tuberculosis (I_T) are the recovery rate of individuals infected with tuberculosis (α_T), which has a negative correlation, and the infection rate of tuberculosis (γ_T),

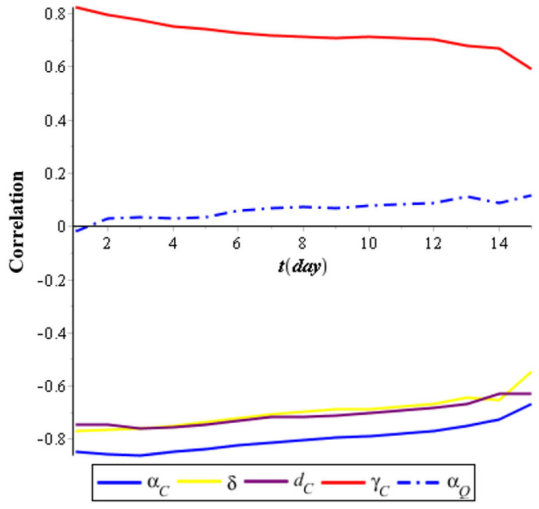


FIGURE 2
PRCC of individuals infected with COVID-19 (I_C).

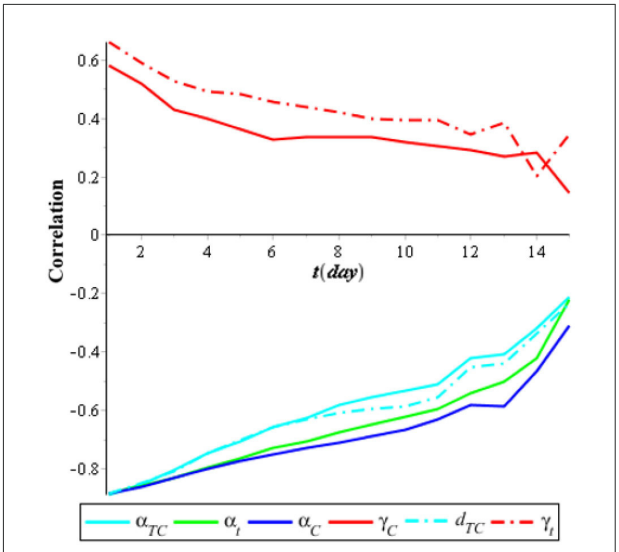


FIGURE 4
PRCC of individuals coinfected with COVID-19 and tuberculosis (I_{TC}).

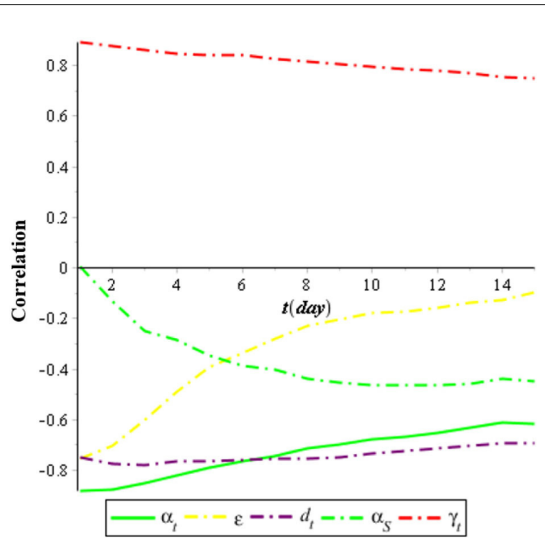


FIGURE 3
PRCC of individuals infected with tuberculosis (I_T).

which has a positive correlation. Meanwhile, the treatment rate also has a negative correlation with the individuals infected with tuberculosis (I_T). It means that the effort to treat individuals infected with tuberculosis (I_T) can suppress the spreading of tuberculosis.

Meanwhile, Figure 4 shows that the most sensitive parameters to individuals coinfected with COVID-19 and tuberculosis (I_{TC}) are $\alpha_C, \alpha_T, and\ \alpha_{TC}$, which has a negative correlation, and γ_C *and* γ_T, which has a positive correlation. It means that to minimize the case of coinfection of COVID-19 and tuberculosis, we need a combination of interventions for COVID-19 and tuberculosis.

Optimal control

The aim of the optimal control is to reduce the spreading of COVID-19 and tuberculosis. In this research, we used prevention control for COVID-19 and tuberculosis, i.e., the use of masks and hand sanitizer. There are two controls in the model. The COVID-19 prevention control $u(t)$ is to minimize the individuals infected with COVID-19 (I_C). The tuberculosis prevention control $v(t)$ is to minimize the individuals infected with tuberculosis (I_T). The objective function is given by the following equation:

$$\min_{(u,v)} J(u,v) = \min_{(u,v)} \int_0^{t_f} \left(C_1 I_C(t) + W_1 u^2(t) + C_2 I_T(t) + W_2 v^2(t) \right) dt \tag{10}$$

where $0 \le u(t) \le 1, 0 \le v(t) \le 1, 0 \le t \le t_f$

$$\begin{aligned} & s.t \\ \frac{dS}{dt} &= A - ((1-u(t))\gamma_C (I_C(t) + I_{TC}(t)) \\ &\quad + (1-v(t))\gamma_T (I_T(t) + I_{TC}(t) + T(t)) \\ &\quad + \mu) S(t) \\ \frac{dI_C}{dt} &= \left(1-u(t)\right)\gamma_C \left(I_C(t) + I_{TC}(t)\right) S(t) + \alpha_T I_{TC}(t) \\ &\quad - (\gamma_T (I_T(t) + I_{TC}(t) + T(t)) + \delta + \mu + \alpha_C \end{aligned} \tag{11}$$

$$+d_C)\, I_C(t) \tag{12}$$

$$\frac{dI_T}{dt} = (1-v(t))\,\gamma_T\,(I_T(t)+I_{TC}(t)+T(t))\,S(t) + \alpha_C I_{TC}(t) - (\gamma_C\,(I_C(t)+I_{TC}(t)) + \varepsilon + \mu + \alpha_T + d_T)\, I_T(t) \tag{13}$$

$$\frac{dI_{TC}}{dt} = \gamma_C\,(I_C(t)+I_{TC}(t))\,I_T(t) + \gamma_T\,(I_T(t)+I_{TC}(t) + T(t))\,I_C(t) - (\alpha_{TC}+\mu+\alpha_T+\alpha_C + d_{TC})\, I_{TC}(t) \tag{14}$$

$$\frac{dQ}{dt} = \delta I_C(t) - (\alpha_Q+\mu+d_C)\,Q(t) \tag{15}$$

$$\frac{dT}{dt} = \varepsilon I_T(t) - (\alpha_S+\mu+d_T)\,T(t) \tag{16}$$

$$\frac{dR}{dt} = \alpha_C I_C(t) + \alpha_T I_T(t) + \alpha_Q Q(t) + \alpha_S T(t) + \alpha_{TC} I_{TC}(t) - \mu R(t), \tag{17}$$

where $S(t) \geq 0, I_C(t) \geq 0, I_T(t) \geq 0, I_{TC}(t) \geq 0, Q(t) \geq 0, T(t) \geq 0, R(t) \geq 0$.

C_1 represents the positive constant weight that balance off the individuals infected with COVID-19 to COVID-19 and C_2 represents the positive constant weight that balance off the individuals infected with tuberculosis to tuberculosis. W_1 and W_2 are the constant weights of the prevention control for COVID-19 and tuberculosis. $W_1u^2(t)$ and $W_2v^2(t)$ represent the costs of the prevention control for COVID-19 and tuberculosis. We used a quadratic control function because the positive balancing cost factors transfer the integral into monetary quantity over a finite period of time [see Tchoumi et al. [45] and the references therein].

The Hamiltonian function for the optimal control model is given by the following equation:

$$\begin{aligned} H = {} & C_1 I_C(t) + W_1 u^2(t) + C_2 I_T(t) + W_2 v^2(t) \\ & + \lambda_1 \big(A - \big((1-u(t))\,\gamma_C\,(I_C(t) + I_{TC}(t)\big) + (1-v(t))\,\gamma_T\,(I_T(t)+I_{TC}(t)+T(t)) + \mu)\, S(t)\big) \\ & + \lambda_2 \big((1-u(t))\,\gamma_C\,(I_C(t)+I_{TC}(t))\,S(t) + \alpha_T I_{TC}(t) - (\gamma_T\,(I_T(t)+I_{TC}(t)+T(t)) + \delta + \mu + \alpha_C + d_C)\, I_C(t)\big) \\ & + \lambda_3 \big((1-v(t))\,\gamma_T\,(I_T(t)+I_{TC}(t)+T(t))\,S(t) + \alpha_C I_{TC}(t) - (\gamma_C\,(I_C(t)+I_{TC}(t)) + \varepsilon + \mu + \alpha_T + d_T)\, I_T(t)\big) \\ & + \lambda_4 \big(\gamma_C\,(I_C(t)+I_{TC}(t))\,I_T(t) + \gamma_T\,(I_T(t)+I_{TC}(t) + T(t))\,I_C(t) - (\alpha_{TC}+\mu+\alpha_T+\alpha_C+d_{TC})\, I_{TC}(t)\big) \\ & + \lambda_5 \big(\delta I_C(t) - (\alpha_Q+\mu+d_C)\,Q(t)\big) \\ & + \lambda_6 \big(\varepsilon I_T(t) - (\alpha_S+\mu+d_T)\,T(t)\big) \\ & + \lambda_7 \big(\alpha_C I_C(t) + \alpha_T I_T(t) + \alpha_Q Q(t) + \alpha_S T(t) + \alpha_{TC} I_{TC}(t) - \mu R(t)\big), \end{aligned}$$

with λ_i for $i = 1, 2, \ldots, 7$ is the adjoint variable of $S(t), I_C(t), I_T(t), I_{TC}(t), Q(t), T(t), R(t)$.

The co-state equation of the optimal control model is given by the following equation:

$$\begin{aligned} \frac{d\lambda_1}{dt} = {} & -\lambda_1(t)\big(-(1-u(t))\,\gamma_C\,(I_C(t)+I_{TC}(t)) - (1-v(t))\,\gamma_T\,(I_T(t)+I_{TC}(t)+T(t)) - \mu\big) \\ & -\lambda_2(t)\big((1-u(t))\,\gamma_C\,(I_C(t)+I_{TC}(t))\big) \\ & -\lambda_3(t)\,(1-v(t))\,\gamma_T\,(I_T(t)+I_{TC}(t)+T(t)) \\ \frac{d\lambda_2}{dt} = {} & -C_1 + \lambda_1(t)\,(1-u(t))\,\gamma_C S(t) \\ & -\lambda_2(t)\,((1-u(t))\,\gamma_C S(t) - (\gamma_T\,(I_T(t)+I_{TC}(t) + T(t)) + \delta + \mu + \alpha_C + d_C)) \\ & +\lambda_3(t)\,\gamma_C I_T(t) - \lambda_4(t)\,(\gamma_C I_T(t) + \gamma_T\,(I_T(t) + I_{TC}(t) + T(t))) - \lambda_5(t)\delta - \lambda_7(t)\alpha_C \\ \frac{d\lambda_3}{dt} = {} & -C_2 + \lambda_1(t)\,(1-v(t))\,\gamma_T S(t) + \lambda_2(t)\,\gamma_T I_C(t) \\ & -\lambda_3(t)\,((1-v(t))\,\gamma_T S(t) - (\gamma_C\,(I_C(t)+I_{TC}(t)) + \varepsilon + \mu + \alpha_T + d_T)) \\ & -\lambda_4(t)\,(\gamma_T I_C(t) + \gamma_C\,(I_C(t)+I_{TC}(t))) - \lambda_6(t)\varepsilon - \lambda_7(t)\alpha_T \\ \frac{d\lambda_4}{dt} = {} & \lambda_1(t)\,((1-u(t))\,\gamma_C + (1-v(t))\,\gamma_T)\,S(t) \\ & -\lambda_2(t)\,((1-u(t))\,\gamma_C S(t) + \alpha_T - \gamma_T I_C(t)) \\ & -\lambda_3(t)\,((1-v(t))\,\gamma_T S(t) + \alpha_C - \gamma_C I_T(t)) \\ & -\lambda_4(t)\,(\gamma_T I_C(t) + \gamma_C I_T(t) - \alpha_{TC} - \mu - \alpha_T - \alpha_C - d_{TC}) - \lambda_7(t)\alpha_{TC} \\ \frac{d\lambda_5}{dt} = {} & -\lambda_5(t)\,(-\alpha_Q - \mu - d_C) - \lambda_5(t)\,\alpha_Q \\ \frac{d\lambda_6}{dt} = {} & \lambda_1(t)\,(1-v(t))\,\gamma_T S(t) + \lambda_2(t)\,\gamma_T I_C(t) \\ & -\lambda_3(t)\,(1-v(t))\,\gamma_T S(t) - \lambda_4(t)\,\gamma_T I_C(t) \\ & -\lambda_6(t)\,(-\alpha_S - \mu - d_T) - \lambda_7(t)\,\alpha_S \\ \frac{d\lambda_7}{dt} = {} & \lambda_7(t)\,\mu. \end{aligned}$$

According to Lenhart and Workman [46], the optimal condition of the optimal control for $0 \leq t \leq t_f$ is given as follows:

$$\frac{\partial H}{\partial u} = 0$$
$$\frac{\partial H}{\partial v} = 0$$

since $0 \leq u(t) \leq 1$ *and* $0 \leq v(t) \leq 1$, we obtained the optimal control as follows:

$$u^*(t) = \min\left\{1, \max\left\{0, \frac{\gamma_C S(t)\,(\lambda_2(t) - \lambda_1(t))\,(I_C(t)+I_{TC}(t))}{2W_1}\right\}\right\}$$

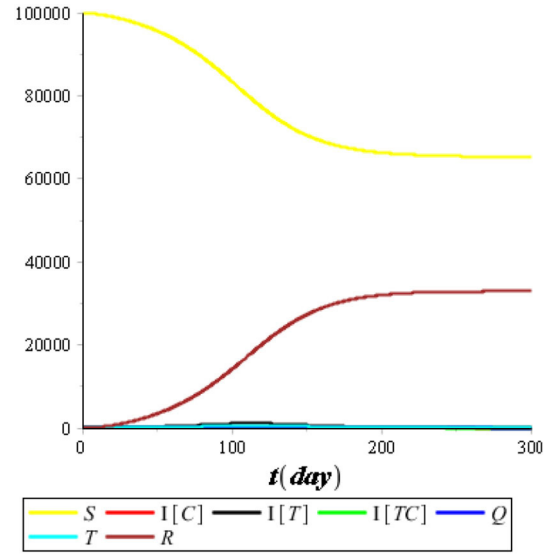


FIGURE 5
The dynamics of all populations.

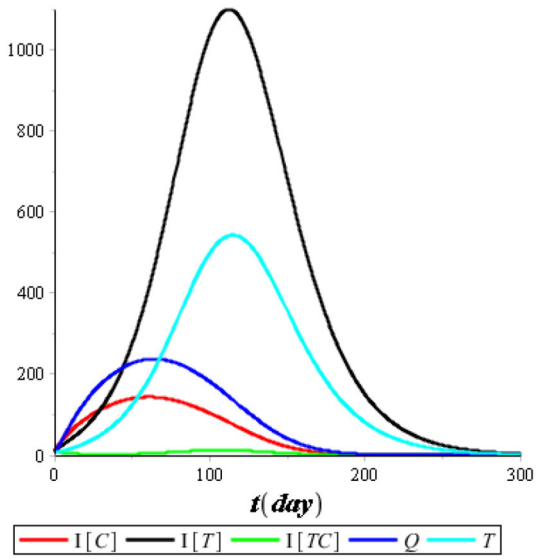


FIGURE 6
The dynamics of the population without susceptible (S) and recovered (R).

$$v^*(t) = \min\left\{1, \max\left\{0, \frac{\gamma_T S(t)\left(\lambda_3(t) - \lambda_1(t)\right)\left(I_T(t) + I_{TC}(t) + T(t)\right)}{2W_2}\right\}\right\}$$

Numerical simulation

The aim of the numerical simulation is to observe the dynamics of the population graphically. The numerical simulation used the compartment's initial values, $\{S(0), I_C(0), I_T(0), I_{TC}(0), Q(0), T(0), R(0)\} = \{100000, 10, 10, 10, 10, 10, 10\}$, and parameter values from Table 2. We performed the numerical simulations for the models with and without control.

Dynamics of the population

First, we performed the numerical simulation without control. The result of the numerical simulation is presented in Figures 5, 6.

Figure 5 illustrates that the recovered population (R) improved considerably compared with the other compartments, so we omitted it from the simulation. Figure 6 illustrates that the most infected individuals are the individuals infected with tuberculosis (I_T), while the individuals infected with COVID-19 (I_C) and the individuals coinfected with COVID-19 and tuberculosis (I_{TC}) are not significant. We conclude that isolation has an immediate impact on reducing the number of COVID-19 infections, while treatment has an impact that tends to take a long time of up to 120 days or more. It happens because people who are undergoing treatment can still transmit tuberculosis as long as the bacteria in their bodies are still active.

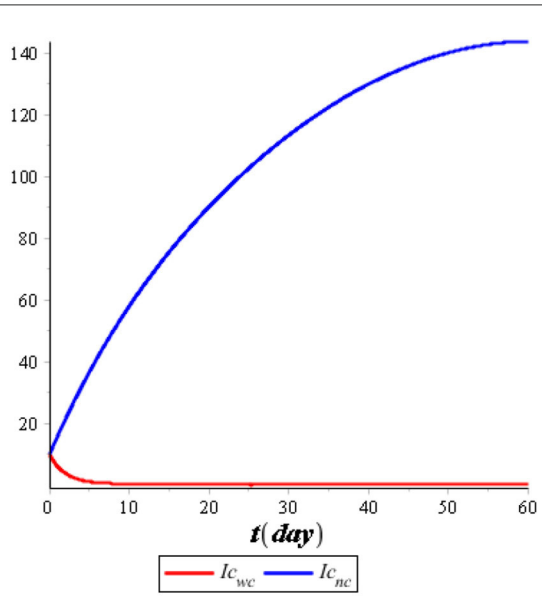


FIGURE 7
The COVID-19-infected individuals with control (Ic_{wc}) and no control (Ic_{nc}).

10.3389/fams.2022.958081

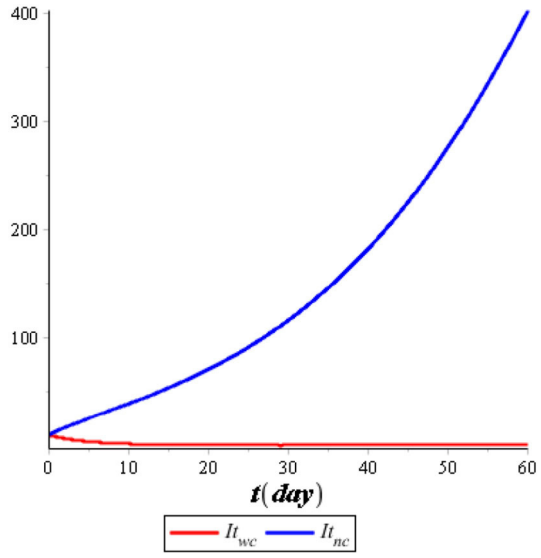


FIGURE 8
The tuberculosis-infected individuals with control (It_{wc}) and no control (It_{nc}).

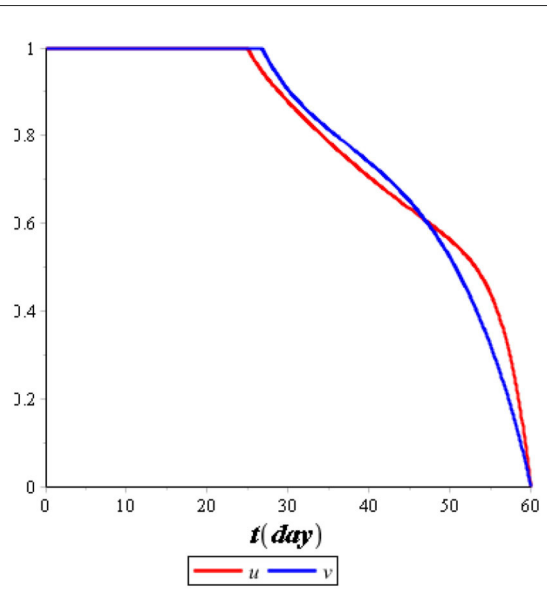


FIGURE 9
Optimal control u^* and v^*.

Optimal control

We performed the numerical simulation using optimal control $(u^*(t), v^*(t))$. The result of the numerical simulation is illustrated in Figures 7, 8, while the optimal control is illustrated in Figure 9.

Figure 7 illustrates the comparison of the dynamics of I_C with control $u^*(t)$ and no control. Also, Figure 8 illustrates the comparison of the dynamics of I_T with control $v^*(t)$ and no control. Both controls are effective in reducing the spreading of COVID-19 and tuberculosis. It means that the prevention control for COVID-19 and tuberculosis is a recommended solution compared with no control for the population. Figure 9 illustrates the most effective control in 60 days. Therefore, we need to use 100% of the prevention control in the first 30 days, which can then be relaxed for the next 30 days.

Conclusion

In this research, we presented a coinfection model of tuberculosis and COVID-19 with the effect of isolation and treatment. The coinfection model has two basic reproduction numbers, namely, the basic reproduction number for COVID-19 and the basic reproduction number for tuberculosis. The disease-free equilibrium is locally asymptotically stable when $\Re_0 < 1$. Otherwise, the endemic equilibrium is locally asymptotically stable when $\Re_0 > 1$. Global sensitivity analysis was performed using the combination of LHS and PRCC. We analyzed the PRCC for three compartments, namely, individuals infected with COVID-19 (I_C), individuals infected with tuberculosis (I_T), and individuals coinfected with COVID-19 and tuberculosis (I_{TC}). The most sensitive parameters to the model are the recovery rate of individuals infected with COVID-19 (α_C), the infection rate of COVID-19 (γ_C), the recovery rate of individuals infected with tuberculosis (α_T), the infection rate of tuberculosis (γ_T), and the recovery rate of individuals coinfected with tuberculosis and COVID-19 (α_{TC}). We performed the optimal control in the form of prevention for COVID-19 and tuberculosis. The numerical simulation shows that these controls effectively reduce the infected population. The effect of isolation has an immediate impact on reducing the number of COVID-19 infections, while the effect of treatment has an impact that tends to take a longer time.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

Author contributions

FI, NA, and AS: study design, conceptualization, and methodology. NA and AS: validation, investigation, and supervision. NA: project administration and funding acquisition. FI: model formulation, software, writing, model

analysis, simulation, and visualization. All authors have read and agreed to the published version of the manuscript.

Funding

This research is funded by Universitas Padjadjaran through Beasiswa Unggulan Pascasarjana Padjadjaran (BUPP) with contract number 2203/UN6.3.1/PT.00/2022.

Acknowledgments

The authors thank the Universitas Padjadjaran for supporting this research. The authors also thank the reviewers for the valuable review of this article.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. United Nations. *Transforming Our World: The 2030 Agenda for Sustainable Development*. https://www.google.com/search?q=San$+$Francisco&stick=H4sIAAAAAAAAAOPgE-LQz9U3MC8xM1UCs1LMcgq0VLOTrfTzi9IT8zKrEksy8_NQOFZp-aV5Kakpi1h5gxPzFNyKEvOSM4uT83ewMu5iZ-JgAAC4aH05VQAAAA&sa=X&ved=2ahUKEwjQ4sui_Yv5AhXiEGIAHbItDfMQmxMoAXoFCI4BEAM San Francisco, CA: United Nations (2015).

2. Riley RL, Mills CC, O'Grady F. Sultan Lu, Wittstadt S, Shuvpuri DN. Infectiousness of air from a tuberculosis ward. *Am Rev Respir Dis.* (1961) 85:511–25.

3. World Health Organization. *World Tuberculosis Day 2022*. Available at: https://www.who.int/campaigns/world-tb-day/2022 (accessed April 30, 2022).

4. Ministry of Health, Indonesia. *Tuberculosis Library Indonesia*. Available at: https://tbindonesia.or.id/pustaka-tbc/dashboard-tb/ (accessed June 24, 2022).

5. Prawoto N, Purnomo EP, Zahra AA. The impacts of Covid-19 pandemic on socio-economic mobility in Indonesia. *Int J Econ Bus Admin.* (2020) 8:57–71. doi: 10.35808/ijeba/486

6. Worldometers. *Reported Cases and Deaths by Country or Territory*. Available at: https://www.worldometers.info/coronavirus/#countries (accessed April 30, 2022).

7. Zegarra MAA, Infante SD, Carrasco DB, Liceaga DO. COVID-19 optimal vaccination policies: A modeling study on efficacy, natural and vaccine-induced immunity responses. *Math Biosci.* (2021) 337:108614. doi: 10.1016/j.mbs.2021.108614

8. Diagne ML, Rwezaura H, Tchoumi SY, Tchuenche JM. A Mathematical Model of COVID-19 with Vaccination and Treatment. *Comput Math Methods Med.* (2021) 2021:250129. doi: 10.1155/2021/1250129

9. Inayaturohmat F, Anggriani N, Supriatna AK. Optimal control and sensitivity analysis of COVID-19 transmission model with the presence of waning immunity in West Java, Indonesia. *Commun Math Biol Neurosci.* (2022) 2022:19. doi: 10.28919/cmbn/7124

10. Inayaturohmat F, Zikkah RN, Supriatna AK, Anggriani N. Mathematical model of COVID-19 transmission in the presence of waning immunity. *J Phys Conf Ser.* (2021) 1722:012038. doi: 10.1088/1742-6596/1722/1/012038

11. Arruda EF, Das SS, Dias CM, Pastore DH. Modelling and optimal control of multi strain epidemics, with application to COVID-19. *PLoS ONE.* (2021) 16:0257512. doi: 10.1371/journal.pone.0257512

12. Rabiu M, Iyaniwura SA. Assessing the potential impact of immunity waning on the dynamics of COVID-19 in South Africa: an endemic model of COVID-19. *Nonlinear Dyn.* (2022) 109:203–23. doi: 10.1007/s11071-022-07225-9

13. Sasmita NR, Ikhwan M, Susyanto S, Chongsuvivatwong V. Optimal control on a mathematical model to pattern the progression of coronavirus disease 2019 (COVID-19) in Indonesia. *Global Health Res Policy.* (2020) 5:38. doi: 10.1186/s41256-020-00163-2

14. Abioye AI, Peter OJ, Ogunseye HA, Oguntolu FA, Oshinubi K, Ibrahim AA, et al. Mathematical model of COVID-19 in Nigeria with optimal control. *Results Phys.* (2021) 28:104598. doi: 10.1016/j.rinp.2021.104598

15. Salman AM, Ahmed I, Mohd MH, Jamiluddin MS, Dheyab, MA. Scenario analysis of COVID-19 transmission dynamics in Malaysia with the possibility of reinfection and limited medical resources scenarios. *Comput Biol Med.* (2021) 133:104372. doi: 10.1016/j.compbiomed.2021.104372

16. Zamir M, Nadeem F, Alqudah MA, Abdeljawad T. Future implications of COVID-19 through Mathematical modelling. *Results Phys.* (2022) 33:105097. doi: 10.1016/j.rinp.2021.105097

17. Masandawa L, Mirau SS, Mbalawata IS. Mathematical modeling of COVID-19 transmission dynamics between healthcare workers and community. *Results Phys.* (2021) 29:104731. doi: 10.1016/j.rinp.2021.104731

18. Kassa SM, Njagarah JBH, Terefe YA. Analysis of the mitigation strategies for COVID-19: From mathematical modelling perspective. *Chaos Solitons Fractals.* (2020) 138:109968. doi: 10.1016/j.chaos.2020.109968

19. Dwomoh D, Iddi S, Adu B, Aheto JM, Sedzro KM, Fobil J, et al. Mathematical modeling of COVID-19 infection dynamics in Ghana: Impact evaluation of integrated government and individual level interventions. *Infect Dis Modell.* (2021) 6:381–97. doi: 10.1016/j.idm.2021.01.008

20. Fatima B, Zaman G. Co-infection of Middle Eastern respiratory syndrome coronavirus and pulmonary tuberculosis. *Chaos Solitons Fractals.* (2020) 140:110205. doi: 10.1016/j.chaos.2020.110205

21. Das P, Nadim SS, Das S, Das P. Dynamics of COVID-19 transmission with comorbidity: a data driven modelling based approach. *Nonlinear Dyn.* (2021) 106:1197–211. doi: 10.1007/s11071-021-06324-3

22. Das P, Upadhyay RK, Misra AK, Rihan FA, Das P, Ghosh D. Mathematical model of COVID-19 with comorbidity and controlling using non-pharmaceutical interventions and vaccination. *Nonlinear Dyn.* (2021) 106:1213–27. doi: 10.1007/s11071-021-06517-w

23. Ssebuliba J, Nakakawa JN, Ssematimba A, Mugisha JYT. Mathematical modelling of COVID-19 transmission dynamics in a partially comorbid community. *Partial Diff Equations Appl Math.* (2022) 5:100212. doi: 10.1016/j.padiff.2021.100212

24. Omame A, Sene N, Nometa I, Nwakanma CI, Nwafor EU, Iheonu N, et al. Analysis of COVID-19 and comorbidity co-infection model with optimal control. *Optimal Control Appl Methods.* (2021) 42:1568–90. doi: 10.1002/oca.2748

25. Agusto FB, Adekunle AI. Optimal control of a two-strain tuberculosis-HIV/AIDS co-infection model. *Biosystems.* (2014) 119:20–44. doi: 10.1016/j.biosystems.2014.03.006

26. Magombedze G, Garira W, Mwenje E. Modeling the TB/HIV-1 Co-infection and the effects of its treatment. *Int J Math Demogr.* (2010) 17:12–64. doi: 10.1080/08898480903467241

27. Fatmawati, TH. An Optimal treatment control of TB-HIV coinfection. *Int J Math Math Sci.* (2016) 2016:8261208. doi: 10.1155/2016/8261208

28. Fatmawati TH. *Optimal control of HIV resistance and tuberculosis co-infection using treatment intervention. Asian Pacific J Trop Dis.* (2017) 7:366–73. doi: 10.12980/apjtd.7.2017D6-400

29. Majumder M, Tiwari PK, Pal S. Impact of saturated treatments on HIV-TB dual epidemic as a consequence of COVID-19: optimal control with awareness and treatment. *Nonlinear Dyn.* (2022) 109:143–76. doi: 10.1007/s11071-022-07395-6

30. Tanvi A, Aggarwal R, Raj YA. A fractional order HIV-TB co-infection model in the presence of exogenous reinfection and recurrent TB. *Nonlinear Dyn.* (2021) 104:4701–25. doi: 10.1007/s11071-021-06518-9

31. Moya EMD, Pietrus A, Oliva SM. Mathematical model with fractional order derivatives for Tuberculosis taking into account its relationship with HIV/AIDS and diabetes. *Jambura J Biomath.* (2021) 2:80–95. doi: 10.34312/jjbm.v2i2.11553

32. Pinto CMA, Carvalho ARM. The HIV/TB coinfection severity in the presence of TB multi-drug resistant strains. *Ecol Complex.* (2017) 32:1–20. doi: 10.1016/j.ecocom.2017.08.001

33. Tahir M, Shah SIA, Zaman G. Prevention strategy for superinfection mathematical model tuberculosis and HIV associated with AIDS. *Cogent Math Stat.* (2019) 6:1637166. doi: 10.1080/25742558.2019.1637166

34. Marimutu Y, Nagappa B, Sharma N, Basu S, Chopra KK. COVID-19 and tuberculosis: A mathematical model based forecasting in Delhi, India. *Indian J Tuberculosis.* (2020) 67:177–81. doi: 10.1016/j.ijtb.2020.05.006

35. Omame A, Abbas M, Onyenegecha CP. A fractional-order model for COVID-19 and tuberculosis co-infection using Atangana–Baleanu derivative. *Chaos Solitons Fractals.* (2021) 153:111486. doi: 10.1016/j.chaos.2021.111486

36. Goudiaby MS, Gning LD, Diagne ML, Dia BM, Rwezaura H, Tchuenche JM. Optimal control analysis of a COVID-19 and tuberculosis co-dynamics model. *Inform Med Unlocked.* (2022) 28:100849. doi: 10.1016/j.imu.2022.100849

37. Mekonen KG, Balcha SF, Obsu LL, Hassen A. Mathematical modeling and analysis of TB and COVID-19 coinfection. *J Appl Math.* (2022) 2022:2449710. doi: 10.1155/2022/2449710

38. Agusto, FB. Optimal isolation control strategies and cost-effectiveness analysis of a two-strain avian influenza model. *Biosystems*. (2013) 113:155–64. doi: 10.1016/j.biosystems.2013.06.004

39. Aissata AB, Mamadou L, Diagne ML, Mbaye I, Seydi O. A mathematical model for the impact of public health education campaign for tuberculosis. *Far East J Appl Math.* (2018) 100:97–138. doi: 10.17654/AM100020097

40. Nicholas JI, Adrian ER. Estimating SARS-CoV-2 infections from deaths, confirmed cases, tests, and random surveys. *Proc Natl Acad Sci (PNAS).* (2021) 118:2103272118. doi: 10.1073/pnas.2103272118

41. Diekmann O, Heesterbeek JAP, Metz JAJ. On the definition and the computation of the basic reproduction ratio R_0 in models for infectious disease in heterogeneous populations. *J Math Biol.* (1990) 28:365–82.

42. Driessche PV, Watmough J. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Math Biosci.* (2002) 180:29–48. doi: 10.1016/S0025-5564(02)00108-6

43. Boyce WE, DiPrima RC. *Elementary Differential Equations and Boundary Value Problems. 9th ed*. New York, NY: John Wiley & Sons, Inc (2009).

44. Marino S, Hogue IB, Ray CJ. A Methodology for performing global uncertainty and sensitivity analysis in systems biology. *J Theor Biol.* (2008) 254:178–96. doi: 10.1016/j.jtbi.2008.04.011

45. Tchoumi SY, Diagne ML, Rwezaura H. Malaria and COVID-19 co-dynamics: A mathematical model and optimal control. *Appl Math Model.* (2021) 99:294–327. doi: 10.1016/j.apm.2021.06.016

46. Lenhart S, Workman JT. *Optimal Control Applied to Biological Models*. Boca Raton: Chapman and Hall/CRC (2007).

10.3389/fams.2022.958081

Appendix 1

Coefficient values of the characteristic polynomial $\mathcal{P}_0$ (λ):

$$k_2 = 1$$
$$k_1 = 0.5990320875$$
$$k_0 = 0.00697110062$$

Appendix 2

Coefficient values of the characteristic polynomial $\mathcal{P}^*(\lambda)$:

$$m_7 = 1$$
$$m_6 = 3.151228668$$
$$m_5 = 3.499353074$$
$$m_4 = 1.853443921$$
$$m_3 = 0.5668236943$$
$$m_2 = 0.1127678992$$
$$m_1 = 0.008741862824$$
$$m_0 = 4.056943389 \times 10^{-7}$$

frontiers | Frontiers in Applied Mathematics and Statistics

TYPE Original Research
PUBLISHED 07 September 2022
DOI 10.3389/fams.2022.953223

OPEN ACCESS

EDITED BY
Mohd Hafiz Mohd,
Universiti Sains Malaysia (USM),
Malaysia

REVIEWED BY
Ummu Atiqah Mohd Roslan,
University of Malaysia Terengganu,
Malaysia
Ojonubah James Omaiye,
Federal College of Education, Okene,
Nigeria

*CORRESPONDENCE
Asep K. Supriatna
a.k.supriatna@unpad.ac.id

SPECIALTY SECTION
This article was submitted to Mathematics of Computation and Data Science, a section of the journal Frontiers in Applied Mathematics and Statistics

RECEIVED 25 May 2022
ACCEPTED 15 August 2022
PUBLISHED 07 September 2022

CITATION
Hasibuan A, Supriatna AK and Carnia E (2022) Local stability analysis of two density-dependent semelparous species in two age classes. *Front. Appl. Math. Stat.* 8:953223. doi: 10.3389/fams.2022.953223

Local stability analysis of two density-dependent semelparous species in two age classes

Arjun Hasibuan[1], Asep K. Supriatna[2]* and Ema Carnia[2]

[1]Master of Mathematics Study Program, Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Sumedang, Indonesia, [2]Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Sumedang, Indonesia

It is crucial to take into account the dynamics of the species while investigating how a species may survive in an environment. A species can be classified as either semelparous or iteroparous depending on how it reproduces. In this article, we present a model, which consists of two semelparous species by considering two age classes. We specifically discuss the effects of density-dependent in the interaction between the two semelparaous species and examine the equilibria of the system in the absence and presence of harvesting in the system. Then, the local stability of the equilibria is also investigated. A modified Leslie matrix population model with the addition of density-dependent in the equation is used. The model is analyzed in the presence and absence of competition between these species. We assume that density-dependent only occurred in the first age class of both species and that harvesting only occurred in the second age class of both species. Then, we assume that competition only occurs in the first age class in both species in the form of interspecific and intraspecific competition. This assumption is intended to simplify the complexity of the problem in the model. Our results show that there are three equilibria in the model without competition and four equilibria in the model with the competition. Hence, the presence of competition has influenced the number of equilibria. We also investigate the relation between the stability of the equilibria with the net reproduction rate of the system. Furthermore, we found the condition for the local stability of the co-existence equilibrium point, which is related to the degree of interspecific and intraspecific competition. This theory may be applied to investigate the dynamics of natural resources, whether in the absence of human exploitation and in the presence of various strategies in managing the exploitation of the resources, such as in fisheries industries.

KEYWORDS

Leslie matrix, semelparous reproduction strategy, population growth model, age structured matrix, co-existence equilibrium, ecological modeling, harvesting, multispecies

1. Introduction

An ecosystem can be occupied by many species, which can be categorized either as semelparous species or iteroparous species. Semelparous species are species that reproduce only once in their life cycle. Some examples of semelparous species include cicada [1, 2], beetles [1], and salmon [3, 4]. On the other hand, iteroparous species are

those that reproduce more than once during their life cycle. Examples of iteroparous species are easier to find than semelparous, including cows, cats, and many more. These species interact, such that the growth of one species is influenced by the growth of the other species. This is due to some limitations in ecosystems, like limited resources and bounded areas of the location. Because of this limitation, the environment may not meet the whole needs of the existing species. The population growth of the species is likely to be affected by their population. Similarly, species harvesting by humans for food may affects the growth of the species.

The dynamics of a population are very important to be studied with the aim of knowing the condition of the population in the future. This is because the growth of a population can lead to the increase, decrease, stability, or extinction of the population. There are several approaches used to describe the dynamics of the growth of a population in one species or multispecies in an ecosystem. One approach among several approaches is to model the growth of one species or multispecies based on population structure. Growth based on population structure focuses on dividing the overall population into smaller population structures. The population structure can be in the form of, among others, age class, stage of development, and population size.

Population structure allows us to study population growth more comprehensively, which cannot be described by other approaches. One example of a population based on a structure is a population based on age class. Population based on age class assumes that each age class has different characteristics including birth and survival rate. A population model based on the structure can be either continuous or discrete time. In recent years, the Leslie matrix model has been widely used as a discrete model of the population growth of a species based on age class. This model was introduced by Leslie [5]. The Leslie matrix model in the linear case form assumes that there are differences in birth rates and survival rates for each age class in population growth. The Leslie matrix model is linear when viewed from its growth graph. This model produces a growth graph that continues to rise, falls continuously, or is constant. Currently, the Leslie model has been widely developed on nonlinear models. The nonlinear Leslie matrix model is more realistic in representing real cases than the linear Leslie matrix model.

The development of the Leslie matrix model is not only limited to its linear and nonlinear forms, but it considers many aspects of the population. For example, the Leslie matrix model initially focused on single species cases, and furthermore, this model has also been developed for multispecies. Some of these studies observed growth models of species whose growth was influenced by density-dependence. Pennycuick et al. [6] focused on simulating the Leslie matrix model on single species and multispecies interactions. Travis et al. [7] conducted a study on Leslie's matrix model for two competing species and simulated the semelparous species. Kon [8] investigated two interacting semelparous species where one species has two age classes, and the other species have one age class. Kon [9] conducted a study on the Leslie matrix model of two semelparous species that have a predator-prey relationship and observed the coprime effect of the number of age classes of the two species. The number of age classes of two species is said to be mutually coprime if the two numbers (the number of classes of two species) have the greatest common factor equal to one. Then, Kon [10] examined the Leslie multispecies semelparous matrix model with an arbitrary number of species and an arbitrary number of interacting age classes. In addition, there are also several multispecies studies using other methods, both continuous time [11, 12], and discrete time [13].

Research on multispecies cases is very important because it is more realistic in describing real cases. However, the multispecies model is more complex than the single species model. This study invests in studying and modeling the population growth of multispecies. Our research focuses on the problem of multispecies population growth, which is influenced by dependence on population density and the effect of harvesting. Of course, this model can be used for both harvested and unharvested species, because if it is not harvested, the harvesting factor is zero. In addition, there are studies on the growth of single species that are affected by harvesting using the Leslie matrix model by Wikan [14]. Then, there is a study on harvesting cases with a different approach by Cooke et al. [15], Getz and Haight [16], Ganguli et al. [17], and Pratama et al. [18] as well as in the multispecies case of Hannesson et al. [19]. Other studies beside the effect of harvesting on the growth of the population, there are also some other studies related to obtaining the maximum sustainable yield, such as Supriatna and Possingham [20], Supriatna [21], Husniah and Supriatna [22], Supriatna and Husniah [23], Supriatna et al. [24], and Husniah et al. [25].

This study is limited to modeling two semelparous species with each species having two age classes. The model used in this paper is the Leslie matrix model for multispecies. In the model, it is assumed that there is a density-dependent influence on the growth of the species which only occurs in the first age class of the two species. In addition, growth for each species is also affected by harvesting which only occurs in the second age class. The model is analyzed under two headings, namely the multispecies model with no effect of interactions and the multispecies model with effect of interactions in the form of interspecific competition (competition between different species) and intraspecific competition (competition between the same species). Competition only occurs in the first age class in both species which is a response due to density-dependence.

This paper is presented as follows. Section 2 discusses the Leslie multispecies matrix model without interactions, determines the equilibrium point, and analyzes the stability

of each equilibrium point. Section 3 discusses the Leslie multispecies matrix model with interactions in the form of interspecific competition and intraspecific competition, determines the equilibrium point of the model, and analyzes the stability of each equilibrium. Section 4 reports on the performance of the simulation, and Section 5 gives the concluding remarks.

2. Leslie matrix model of two semelparous species with density-dependent growth and harvesting without competition

This paper constructs and studies the growth dynamics of two semelparous species with each species having two age classes using the Leslie matrix for multispecies. The model with the number of two age classes can be applied to species that have a lifespan of 2 weeks, 2 months, 2 years, and two age stages can be children and adults, and so on. The model in this section is modified and constructed from the Travis et al. [7] and Wikan [14] models based on the assumptions and simplifications in this paper.

It is assumed that density-dependence only occurs in the first age class in each species. Then, harvesting only occurred in the second age class for both species. Both density-dependent and harvesting affect the population size of the second age class at time $t+1$. In this case, the density-dependent factor uses the classical Beverton-Holt function whose effect is shown in Figure 1. The Beverton-Holt function is also used in research using the Leslie matrix by Wikan [14], but in this study, there is a simplification of the Beverton-Holt function. This model is given by the system of Equations (1) and is called Model A.

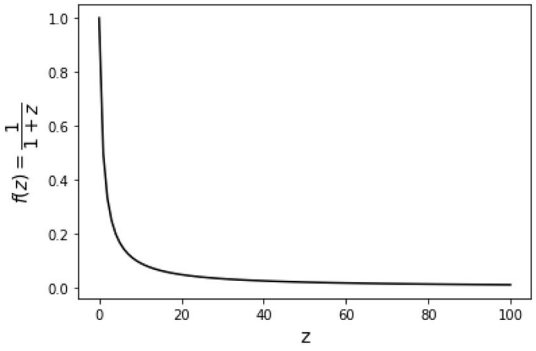


FIGURE 1
Density-dependent Beverton-Holt function.

$$\begin{aligned} x_1(t+1) &= f_{x_2}x_2(t),\\ x_2(t+1) &= \frac{s_{x_1}(1-h_{x_2})}{1+x_1(t)+y_1(t)}x_1(t),\\ y_1(t+1) &= f_{y_2}y_2(t),\\ y_2(t+1) &= \frac{s_{y_1}(1-h_{y_2})}{1+x_1(t)+y_1(t)}y_1(t), \end{aligned} \tag{1}$$

where $x_i(t)$ and $y_i(t)$ represent the population densities of semelparous species x and y from the i-th age class ($i=1,2$) at time t, respectively. In this case $f_{x_2}, f_{y_2} > 0$ represent the birth rate of species x and y in the second age class, respectively. Meanwhile, $0 < s_{x_1}, s_{y_1} \leq 1$ each represents the survival rate of the first age to the second age class of species x and y. Finally, $0 < h_{x_2}, h_{y_2} \leq 1$ are the harvesting rate in the second age class of species x and y, respectively. Based on the assumption, the density-dependent effect in Model A is modeled as $\frac{1}{1+x_1(t)+y_1(t)}$ where z in Figure 1 is $x_1(t)+y_1(t)$. This is based on the assumption that density-dependence only occurs in the first age class in each species.

The equilibrium point of model A can be done by making the left-hand side of model A dependent on time t so that the equilibria are given by the following system of equations:

$$\begin{aligned} x_1(t) &= f_{x_2}x_2(t),\\ x_2(t) &= \frac{s_{x_1}(1-h_{x_2})}{1+x_1(t)+y_1(t)}x_1(t),\\ y_1(t) &= f_{y_2}y_2(t),\\ y_2(t) &= \frac{s_{y_1}(1-h_{y_2})}{1+x_1(t)+y_1(t)}y_1(t). \end{aligned} \tag{2}$$

The system has the following equilibria:

i. The extinction equilibrium point

$$E_0 = \begin{bmatrix} x_1\\ x_2\\ y_1\\ y_2 \end{bmatrix} = \begin{bmatrix} 0\\ 0\\ 0\\ 0 \end{bmatrix},$$

implies, the extinction of both species.

ii. The extinction of the second species (species y) equilibrium point

$$E_x = \begin{bmatrix} x_1\\ x_2\\ y_1\\ y_2 \end{bmatrix} = \begin{bmatrix} (-f_{x_2}s_{x_1}(h_{x_2}-1)-1)\\ -\frac{(f_{x_2}s_{x_1}(h_{x_2}-1)+1)}{f_{x_2}}\\ 0\\ 0 \end{bmatrix} = \begin{bmatrix} R_x-1\\ -\frac{(-R_x+1)}{f_{x_2}}\\ 0\\ 0 \end{bmatrix},$$

with $R_x = f_{x_2}s_{x_1}(1-h_{x_2})$, where R_x represents the expected number of offspring per individual per lifetime when density-dependent effects are neglected on harvest-influenced growth of species x. Therefore, E_x exists if $R_x > 1$.

iii. The extinction of the first species (species x) equilibrium point

$$E_y = \begin{bmatrix} x_1 \\ x_2 \\ y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ (-f_{y_2} s_{y_1}(h_{y_2} - 1) - 1) \\ -\frac{(f_{y_2} s_{y_1}(h_{y_2} - 1) + 1)}{f_{y_2}} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ R_y - 1 \\ -\frac{(-R_y + 1)}{f_{y_2}} \end{bmatrix},$$

where $R_y = f_{y_2} s_{y_1}(1 - h_{y_2})$. R_y represents the expected number of offspring per individual per lifetime when density-dependent effects are neglected on harvest-influenced growth of species y. Therefore, E_y exists if $R_y > 1$.

Based on the calculations, it is not interesting that the point of equilibrium at which both populations exist is not found in this model. Therefore, this model will be developed in Section 3 by adding interaction factors between species.

The following steps analyze the local stability of Model A. An equilibrium point is said to be asymptotically locally stable if the spectral radius of the Jacobian matrix at that equilibrium point is less than one. This criterion is difficult to apply to this model, so in this study, we use the M-Matrix method as done by Travis et al. [7]. The steps taken in the study of Travis et al. [7] to determine local stability asymptotically using the M-Matrix theory, include

i. Determining the Jacobian matrix from the equilibrium point. Suppose $J(E)$ is an $n \times n$ matrix with E being the equilibrium point.
ii. Transformimg the $J(E)$ matrix into a $G = I_n - SJ(E)S^{-1}$ matrix with I_n identity matrix of size $n \times n$ with S details can be seen in Travis et al. [7].
iii. Furthermore, the G matrix has a spectral radius of less than one if G is a M-matrix. A matrix is said to be M-Matrix if $g_{ij} \leq 0$ for $i \neq j$ and if matrix G satisfies one of the five conditions, one of the conditions are that the principal minor of matrix G is positive [see for other conditions in Travis et al. [7]].

The locally stable asymptotic of each equilibrium point of Model A is presented in Theorem 1 below.

Theorem 1. *For systems of Model A:*

i. The equilibrium point E_0 is locally stable asymptotically if $R_x < 1$ and $R_y < 1$.
ii. The equilibrium point E_x is locally stable asymptotically if $R_y < R_x$ and $R_x > 1$.
iii. The equilibrium point E_y is locally stable asymptotically if $R_x < R_y$ and $R_y > 1$.

Proof. The local stability of each equilibrium point can be determined by linearizing the Model A. The Jacobian matrix of Model A becomes

$$J([x_1, x_2, y_1, y_2]) = \begin{bmatrix} 0 & f_{x_2} & 0 & 0 \\ A_x(1 + y_1) & 0 & -A_x x_1 & 0 \\ 0 & 0 & 0 & f_{y_2} \\ -A_y y_1 & 0 & A_y(1 + x_1) & 0 \end{bmatrix}, \quad (3)$$

where

$$A_x = \frac{s_{x_1}(1 - h_{x_2})}{(1 + x_1 + y_1)^2},$$
$$A_y = \frac{s_{y_1}(1 - h_{y_2})}{(1 + x_1 + y_1)^2}.$$

The next step is to substitute each equilibrium point into (3).

i. The Jacobian matrix for the equilibrium point E_0 is obtained as follows.

$$J(E_0) = \begin{bmatrix} 0 & f_{x_2} & 0 & 0 \\ s_{x_1}(1 - h_{x_2}) & 0 & 0 & 0 \\ 0 & 0 & 0 & f_{y_2} \\ 0 & 0 & s_{y_1}(1 - h_{y_2}) & 0 \end{bmatrix}.$$

Based on the method used by Travis et al. [7], the Jacobian matrix $J(E_0)$ will be transformed into a G matrix where $G = I_4 - SJ(E_0)S^{-1}$. Because the value of $J(E_0)_{13}, J(E_0)_{14}, J(E_0)_{23}, J(E_0)_{24}, J(E_0)_{31}, J(E_0)_{32}, J(E_0)_{33}, J(E_0)_{34} \leq 0$ so that the chosen matrix S, i.e.,

$$S = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}.$$

Therefore,

$$G = I_4 - SJ(E_0)S^{-1} = \begin{bmatrix} 1 & -f_{x_2} & 0 & 0 \\ -s_{x_1}(1 - h_{x_2}) & 1 & 0 & 0 \\ 0 & 0 & 1 & -f_{y_2} \\ 0 & 0 & -s_{y_1}(1 - h_{y_2}) & 1 \end{bmatrix}.$$

Based on the defined parameters, the value of $g_{ij} \leq 0$ for $i \neq j$ is fulfilled, where g_{ij} represents the elements of the G matrix in the i-th row of the j-th column so that G is called the M-matrix. Furthermore, the equilibrium point E_0 is locally stable asymptotically if all the minor principals of G are positive. Based on the calculation results obtained

$$\begin{aligned} |g_{11}| &= 1 > 0, \\ \begin{vmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{vmatrix} &= 1 + f_{x_2} s_{x_1}(h_{x_2} - 1) \\ &= 1 - R_x > 0 \implies R_x < 1, \end{aligned}$$

$$\begin{vmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{vmatrix} = 1 + f_{x_2} s_{x_1} (h_{x_2} - 1)$$
$$= 1 - R_x > 0 \implies R_x < 1,$$
$$|G| = (1 + f_{x_2} s_{x_1} (h_{x_2} - 1))(1 + f_{y_2} s_{y_1} (h_{y_2} - 1)) = (1 - R_x)(1 - R_y) > 0,$$
$$\implies R_x < 1 \text{ and } R_y < 1.$$

So, all the minor principals of G will be positive if $R_x < 1$ and $R_y < 1$. Therefore, the equilibrium point E_0 is locally stable asymptotically if $R_x < 1$ and $R_y < 1$.

ii. The Jacobian matrix for the equilibrium point E_x, i.e.,

$$J(E_x) = \begin{bmatrix} 0 & f_{x_2} & 0 & 0 \\ \frac{s_{x_1}(1-h_{x_2})}{R_x^2} & 0 & -\frac{s_{x_1}(1-h_{x_2})(R_x-1)}{R_x^2} & 0 \\ 0 & 0 & 0 & f_{y_2} \\ 0 & 0 & -\frac{s_{y_1}(h_{y_2}-1)}{R_x} & 0 \end{bmatrix}.$$

Next, the Jacobian matrix $J(E_x)$ is transformed into a matrix $G = I_4 - SJ(E_x)S^{-1}$. Because $R_x > 1$ and $J(E_x)_{13}, J(E_x)_{14}, J(E_x)_{23}, J(E_x)_{24}, J(E_x)_{31}, J(E_x)_{32}, J(E_x)_{33}$, and $J(E_x)_{34} \leq 0$ then,

$$S = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}.$$

Furthermore,

$$G = \begin{bmatrix} 1 & -f_{x_2} & 0 & 0 \\ \frac{-s_{x_1}(1-h_{x_2})}{R_x^2} & 1 & \frac{-s_{x_1}(1-h_{x_2})(R_x-1)}{R_x^2} & 0 \\ 0 & 0 & 1 & -f_{y_2} \\ 0 & 0 & \frac{-s_{y_1}(1-h_{y_2})}{R_x} & 1 \end{bmatrix}.$$

Because $R_x > 1$ and the parameters are already defined, $g_{ij} \leq 0$ is fulfilled for $i \neq j$. The next step is to show that the principal minor of G is positive.

Based on the calculation results obtained

$$|g_{11}| = 1 > 0,$$
$$\begin{vmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{vmatrix} = \frac{1 + f_{x_2} s_{x_1} (h_{x_2} - 1)}{f_{x_2} s_{x_1} (h_{x_2} - 1)}$$
$$= -\frac{1 - R_x}{R_x} > 0 \implies R_x > 1,$$
$$\begin{vmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{vmatrix} = \frac{1 + f_{x_2} s_{x_1} (h_{x_2} - 1)}{f_{x_2} s_{x_1} (h_{x_2} - 1)}$$
$$= -\frac{1 - R_x}{R_x} > 0 \implies R_x > 1,$$
$$|G| = \frac{(R_y - R_x)(1 - R_x)}{R_x^2} > 0$$
$$\implies R_x > 1 \text{ and } R_y < R_x.$$

So, all minor principals of G will be positive if $R_x > 1$ and $R_y < R_x$. Therefore, the equilibrium point E_x is locally stable asymptotically if $R_x > 1$ and $R_y < R_x$.

iii. The Jacobian matrix for the equilibrium point E_y is

$$J(E_y) = \begin{bmatrix} 0 & f_{x_2} & 0 & 0 \\ \frac{s_{x_1}(1-h_{x_2})}{R_y} & 0 & 0 & 0 \\ 0 & 0 & 0 & f_{y_2} \\ \frac{-s_{y_1}(1-h_{y_2})(R_y-1)}{R_y^2} & 0 & -\frac{s_{y_1}(h_{y_2}-1)}{R_y} & 0 \end{bmatrix}.$$

Next, the Jacobian matrix $J(E_y)$ is transformed into a matrix $G = I_4 - SJ(E_y)S^{-1}$. Because $R_y > 1$ and $J(E_y)_{13}, J(E_y)_{14}, J(E_y)_{23}, J(E_y)_{24}, J(E_y)_{31}, J(E_y)_{32}, J(E_y)_{33}$, $J(E_y)_{34} \leq 0$ then,

$$S = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}.$$

Furthermore,

$$G = \begin{bmatrix} 1 & -f_{x_2} & 0 & 0 \\ \frac{-s_{x_1}(1-h_{x_2})}{R_y} & 1 & 0 & 0 \\ 0 & 0 & 1 & -f_{y_2} \\ \frac{-s_{y_1}(1-h_{y_2})(R_y-1)}{R_y^2} & 0 & \frac{-s_{y_1}(1-h_{y_2})}{R_y^2} & 1 \end{bmatrix}.$$

Because $R_y > 1$ and the parameters are already defined, $g_{ij} \leq 0$ is fulfilled for $i \neq j$. The next step is to show that the principal minor of G is positive. Based on the calculation results obtained

$$|g_{11}| = 1 > 0,$$
$$\begin{vmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{vmatrix} = \frac{-R_x + R_y}{R_y} > 0 = \frac{-R_x}{R_y} + 1 > 0$$
$$\implies \frac{R_x}{R_y} < 1 \implies R_x < R_y,$$
$$\begin{vmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{vmatrix} = \frac{-R_x + R_y}{R_y} > 0 = \frac{-R_x}{R_y} + 1 > 0$$
$$\implies \frac{R_x}{R_y} < 1 \implies R_x < R_y,$$
$$|G| = -\frac{(R_y - R_x)(1 - R_y)}{R_y^2} > 0$$
$$\implies R_y > 1 \text{ and } R_x < R_y.$$

So, all minor principals of G will be positive if $R_y > 1$ and $R_x < R_y$. Therefore, the equilibrium point E_y is locally stable asymptotically if $R_y > 1$ and $R_x < R_y$. This proof is complete. ■

10.3389/fams.2022.953223

3. Leslie matrix model of two semelparous species with density-dependent growth and harvesting together with competition effect

Based on the previous model, there is no equilibrium point where the two species exist. In fact, in an ecosystem, it is expected that both species can survive. In this section, the previous model is improved upon. An assumption is added to the model that there are both intraspecific and interspecific competitions that occur only in the first age class as a density-dependent response. To simplify the problem, it is assumed that the level of competition between the first age class in species x and species y has the same value, i.e., $a > 0$. Then, the level of competition between the first age class in species y against species x and vice versa has the same value, i.e., $b > 0$. This problem is modeled in Equation (4) below and we call it Model B.

$$\begin{aligned} x_1(t+1) &= f_{x_2}x_2(t), \\ x_2(t+1) &= \frac{s_{x_1}(1-h_{x_2})}{1+ax_1(t)+by_1(t)}x_1(t), \\ y_1(t+1) &= f_{y_2}y_2(t), \\ y_2(t+1) &= \frac{s_{y_1}(1-h_{y_2})}{1+bx_1(t)+ay_1(t)}y_1(t). \end{aligned} \tag{4}$$

Model B is an observed model from the model developed by Travis et al. [7] where in this model there is no effect of harvesting. Then, the model from Travis et al. [7] was modified and constructed based on the assumptions and simplifications in this research.

The equilibrium point of Model B can be determined by making the left-hand side of Model B dependent on the t-th time. The equilibrium solution of Model B is expressed in the following system of equations:

$$\begin{aligned} x_1(t) &= f_{x_2}x_2(t), \\ x_2(t) &= \frac{s_{x_1}(1-h_{x_2})}{1+ax_1(t)+by_1(t)}x_1(t), \\ y_1(t) &= f_{y_2}y_2(t), \\ y_2(t) &= \frac{s_{y_1}(1-h_{y_2})}{1+bx_1(t)+ay_1(t)}y_1(t). \end{aligned} \tag{5}$$

By solving the Equation (5), we arrived at the following results:

i. The extinction equilibrium point is given as

$$E_0 = \begin{bmatrix} x_1 \\ x_2 \\ y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

ii. The equilibrium point with species x exists is given as

$$E_x = \begin{bmatrix} x_1 \\ x_2 \\ y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} -\frac{-R_x+1}{a} \\ -\frac{-R_x+1}{af_{x_2}} \\ 0 \\ 0 \end{bmatrix},$$

where $R_x = f_{x_2}s_{x_1}(1-h_{x_2})$. Therefore, E_x exists only if $R_x > 1$.

iii. The equilibrium point with species y exists is given as

$$E_y = \begin{bmatrix} x_1 \\ x_2 \\ y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ -\frac{-R_y+1}{a} \\ -\frac{-R_y+1}{af_{y_2}} \end{bmatrix},$$

where $R_y = f_{y_2}s_{y_1}(1-h_{y_2})$. Hence, E_y exists only if $R_y > 1$.

iv. The equilibrium point with both species exist given as

$$E_{xy} = \begin{bmatrix} x_1 \\ x_2 \\ y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} -\frac{(a(-R_x+1)-b(-R_y+1))}{(a^2-b^2)} \\ -\frac{(a(-R_x+1)-b(-R_y+1))}{(a^2-b^2)f_{x_2}} \\ -\frac{(a(-R_y+1)-b(-R_x+1))}{(a^2-b^2)} \\ -\frac{(a(-R_y+1)-b(-R_x+1))}{(a^2-b^2)f_{y_2}} \end{bmatrix}.$$

There are two cases where the elements in E_{xy} are positive, namely

Case 1 : $(a^2-b^2) > 0, a(-R_x+1)-b(-R_y+1) < 0$, and $a(-R_y+1)-b(-R_x+1) < 0$.

Case 2 : $(a^2-b^2) < 0, a(-R_x+1)-b(-R_y+1) > 0$, and $a(-R_y+1)-b(-R_x+1) > 0$.

Next, an asymptotically local stability analysis in Model A is performed. The steps taken to check the local stability in Model B are the same as in Model B. However, in Model B there are four equilibrium points that must be analyzed while in Model A there are only three equilibrium points due to the absence of a co-existence equilibrium point. The locally stable asymptotic for each equilibrium point of the model in Model B is presented in Theorem 2 below.

Theorem 2. *For systems of Model B:*

i. If $R_x < 1$ and $R_y < 1$, then the equilibrium point E_0 is locally stable asymptotically.

ii. If $R_x > 1$ and $a(1-R_y) > b(1-R_x)$, then the equilibrium point E_x is locally stable asymptotically.

iii. If $R_y > 1$ and $a(1-R_x) > b(1-R_y)$, then the equilibrium point E_y is locally stable asymptotically.

iv. If $(a^2-b^2) > 0, a(1-R_x) < b(1-R_y)$, and $a(1-R_y) < b(1-R_x)$, then the equilibrium point E_{xy} is locally stable asymptotically.

Proof: The Jacobian matrix of Model B is

$$J([x_1,x_2,y_1,y_2]) = \begin{bmatrix} 0 & f_{x_2} & 0 & 0 \\ A_x(1+by_1) & 0 & -A_x b x_1 & 0 \\ 0 & 0 & 0 & f_{y_2} \\ -A_y b y_1 & 0 & A_y(1+bx_1) & 0 \end{bmatrix}, \quad (6)$$

where

$$A_x = \frac{s_{x_1}(1-h_{x_2}}{(ax_1+by_1+1)^2},$$
$$A_y = \frac{s_{y_1}(1-h_{y_2}}{(ax_1+by_1+1)^2}.$$

The next step is to substitute each equilibrium point into the Jacobian matrix in Equation (6).

i. The Jacobian matrix for the equilibrium point E_0, i.e.,

$$J(E_0) = \begin{bmatrix} 0 & f_{x_2} & 0 & 0 \\ s_{x_1}(1-h_{x_2}) & 0 & 0 & 0 \\ 0 & 0 & 0 & f_{y_2} \\ 0 & 0 & s_{y_1}(1-h_{y_2}) & 0 \end{bmatrix}.$$

The same thing was done in determining the locally stable asymptotically of the Model A. First, the Jacobian matrix $J(E_0)$ is transformed into the matrix $G = I_4 - SJ(E_0)S^{-1}$. Because the value of $J(E_0)_{13}, J(E_0)_{14}, J(E_0)_{23}, J(E_0)_{24}, J(E_0)_{31}, J(E_0)_{32}, J(E_0)_{33}, J(E_0)_{34} \leq 0$ so that the chosen matrix S, i.e.,

$$S = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}.$$

Furthermore,

$$G = \begin{bmatrix} 1 & -f_{x_2} & 0 & 0 \\ -s_{x_1}(1-h_{x_2}) & 1 & 0 & 0 \\ 0 & 0 & 1 & -f_{y_2} \\ 0 & 0 & -s_{y_1}(1-h_{y_2}) & 1 \end{bmatrix}.$$

and $g_{ij} \leq 0$ for $i \neq j$. The next step is to show that the principal minor of G is positive. Based on the calculation results obtained

$$\begin{aligned} |g_{11}| &= 1 > 0, \\ \begin{vmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{vmatrix} &= 1 - R_x > 0 \implies R_x < 1, \\ \begin{vmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{vmatrix} &= 1 - R_x > 0 \implies R_x < 1, \\ |G| &= (1-R_x)(1-R_y) > 0 \\ &\implies R_x < 1 \text{ and } R_y < 1. \end{aligned}$$

So, all minor principals of G will be positive if $R_y < 1$ and $R_x < 1$. Therefore, the equilibrium point E_0 is locally stable asymptotically if $R_x < 1$ and $R_y < 1$.

ii. The Jacobian matrix for the equilibrium point E_x, i.e.,

$$J(E_x) = \begin{bmatrix} 0 & f_{x_2} & 0 & 0 \\ \frac{s_{x_1}(1-h_{x_2})}{R_x^2} & 0 & -\frac{bs_{x_1}(1-h_{x_2})(R_x-1)}{aR_x^2} & 0 \\ 0 & 0 & 0 & f_{y_2} \\ 0 & 0 & -\frac{as_{y_1}(h_{y_2}-1)}{(R_x-1)b+a} & 0 \end{bmatrix}.$$

Next, the matrix $J(E_x)$ is transformed into a matrix $G = I_4 - SJ(E_x)S^{-1}$. Since $R_x > 1$, and the value of $J(E_x)_{13}, J(E_x)_{14}, J(E_x)_{23}, J(E_x)_{24}, J(E_x)_{31}, J(E_x)_{32}, J(E_x)_{33}, J(E_x)_{34} \leq 0$ so that the chosen matrix S, i.e.,

$$S = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}.$$

Furthermore,

$$G = \begin{bmatrix} 1 & -f_{x_2} & 0 & 0 \\ -\frac{s_{x_1}(1-h_{x_2})}{R_x^2} & 1 & -\frac{bs_{x_1}(1-h_{x_2})(R_x-1)}{aR_x^2} & 0 \\ 0 & 0 & 1 & -f_{y_2} \\ 0 & 0 & -\frac{as_{y_1}(1-h_{y_2})}{(R_x-1)b+a} & 1 \end{bmatrix}.$$

and $g_{ij} \leq 0$ for $i \neq j$. The next step is to show that the principal minor of G is positive. Based on the calculation results obtained

$$\begin{aligned} |g_{11}| &= 1 > 0, \\ \begin{vmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{vmatrix} &= \frac{R_x - 1}{R_x} > 0 \implies R_x > 1, \\ \begin{vmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{vmatrix} &= \frac{R_x - 1}{R_x} > 0 \implies R_x > 1, \\ |G| &= \frac{(R_x-1)(a(1-R_y)-b(1-R_x))}{R_x(a-b(1-R_x))}. \end{aligned}$$

For $|G|$, $R_x - 1 > 0$ and $(a - b(1-R_x)) > 0$ because $R_x > 1$. Therefore, for $|G| > 0$, it must be $(a(1-R_y) - b(1-R_x)) > 0$ so that $a(1-R_y) > b(1-R_x)$. Therefore, the equilibrium point E_x is locally stable asymptotically if $R_x > 1$ and $a(1-R_y) > b(1-R_x)$.

iii. The Jacobian matrix for the equilibrium point E_y, i.e.,

$$J(E_y) = \begin{bmatrix} 0 & f_{x_2} & 0 & 0 \\ \frac{as_{x_1}(1-h_{x_2})}{(R_y-1)b+a} & 0 & 0 & 0 \\ 0 & 0 & 0 & f_{y_2} \\ -\frac{bs_{y_1}(1-h_{y_2})(R_y-1)}{aR_y^2} & 0 & -\frac{s_{y_1}(h_{y_2}-1)}{(R_y^2} & 0 \end{bmatrix}.$$

Next, the matrix $J(E_x)$ is transformed into a matrix $G = I_4 - SJ(E_y)S^{-1}$. Since $R_y > 1$, the value of $J(E_y)_{13}, J(E_y)_{14}, J(E_y)_{23}, J(E_y)_{24}, J(E_y)_{31}, J(E_y)_{32}, J(E_y)_{33}$, and $J(E_y)_{34} \leq 0$ so that the chosen matrix S, i.e.,

$$S = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}.$$

Furthermore,

$$G = \begin{bmatrix} 1 & -f_{x_2} & 0 & 0 \\ -\frac{as_{x_1}(1-h_{x_2})}{(R_y-1)b+a} & 1 & 0 & 0 \\ 0 & 0 & 1 & -f_{y_2} \\ -\frac{bs_{y_1}(1-h_{y_2})(R_y-1)}{aR_y^2} & 0 & -\frac{s_{y_1}(1-h_{y_2})}{(R_y^2} & 1 \end{bmatrix}.$$

and $g_{ij} \leq 0$ for $i \neq j$. The next step is to show that the principal minor of G is positive. Based on the calculation results obtained

$$\begin{aligned} |g_{11}| &= 1 > 0, \\ \begin{vmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{vmatrix} &= \frac{(a(1-R_x) - b(1-R_y))}{a - b(1-R_y)}, \\ \begin{vmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{vmatrix} &= \frac{(a(1-R_x) - b(1-R_y))}{a - b(1-R_y)}, \\ |G| &= -\frac{(1-R_y)(a(1-R_x) - b(1-R_y))}{R_y(a - b(1-R_y))}. \end{aligned}$$

Since $R_y > 1$, consequently, $(1 - R_y) < 0$ and $(a - b(1 - R_y)) > 0$. Therefore, for all minor principals to be positive, it must be $(a(1 - R_x) - b(1 - R_y)) > 0$ so that $a(1 - R_x) > b(1 - R_y)$. Thus, the equilibrium point E_y is locally stable asymptotically if $R_y > 1$ and $a(1 - R_x) > b(1 - R_y)$.

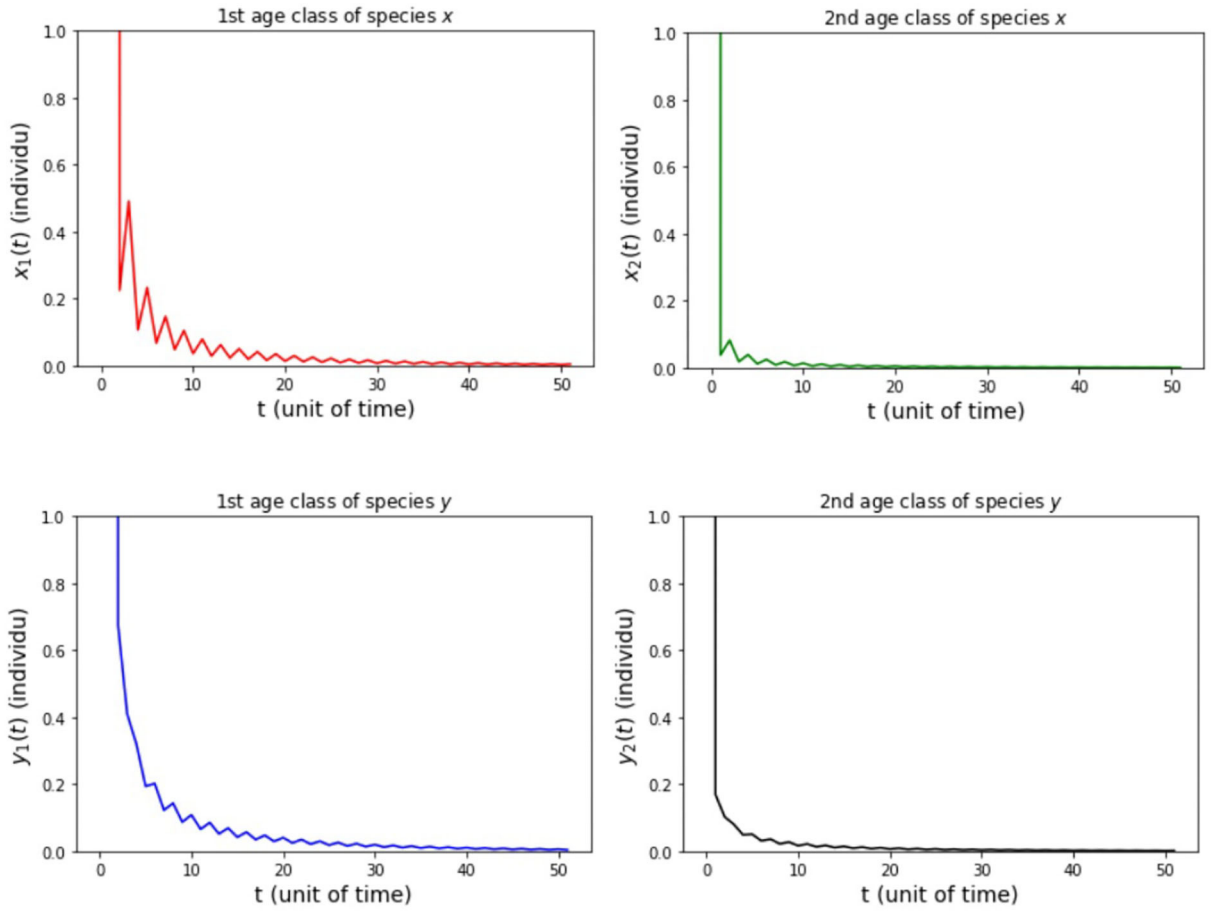


FIGURE 2
Population growth of each age class from case (i) Model A.

iv. The Jacobian matrix for the equilibrium point E_{xy}, i.e.,

$$J(E_{xy}) = \begin{bmatrix} 0 & f_{x_2} & 0 & 0 \\ A_1 & 0 & A_2 & 0 \\ 0 & 0 & 0 & f_{y_2} \\ B_1 & 0 & B_2 & 0 \end{bmatrix},$$

where

$$\begin{aligned} A_1 &= \frac{b^2 R_x + a(-a + (1 - R_y)b)}{-R_x f_{x_2}(a^2 - b^2)}, \\ A_2 &= \frac{b(a(1 - R_x) - b(1 - R_y))}{R_x f_{x_2}(a^2 - b^2)}, \\ B_1 &= \frac{b(a(1 - R_y) - b(1 - R_x))}{R_y f_{y_2}(a^2 - b^2)}, \\ B_2 &= \frac{b^2 R_y + a(-a + (1 - R_x)b)}{-R_y f_{y_2}(a^2 - b^2)}. \end{aligned}$$

Next, the matrix $J(E_{xy})$ is transformed into a matrix $G = I_4 - SJ(E_{xy})S^{-1}$. Because between $a(1-R_x)-b(1-R_y)$ and $a(1 - R_y) - b(1 - R_x)$ to $a^2 - b^2$ have different signs which is a condition where the equilibrium point E_{xy} exists. As a result, A_2 and B_2 are negative. Therefore, the value of $J(E_{xy})_{13}, J(E_{xy})_{14}, J(E_{xy})_{23}, J(E_{xy})_{24}, J(E_{xy})_{31}, J(E_{xy})_{32}$, $J(E_{xy})_{33}$, and $J(E_{xy})_{34} \leq 0$ so that the chosen matrix S, i.e.,

$$S = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}.$$

Furthermore,

$$G = \begin{bmatrix} 1 & -f_{x_2} & 0 & 0 \\ A_1 & 1 & A_2 & 0 \\ 0 & 0 & 1 & -f_{y_2} \\ B_1 & 0 & B_2 & 1 \end{bmatrix}.$$

In addition, $g_{ij} \leq 0$ for $i \neq j$. The next step is to show that the principal minor of G is positive. Based on the calculation

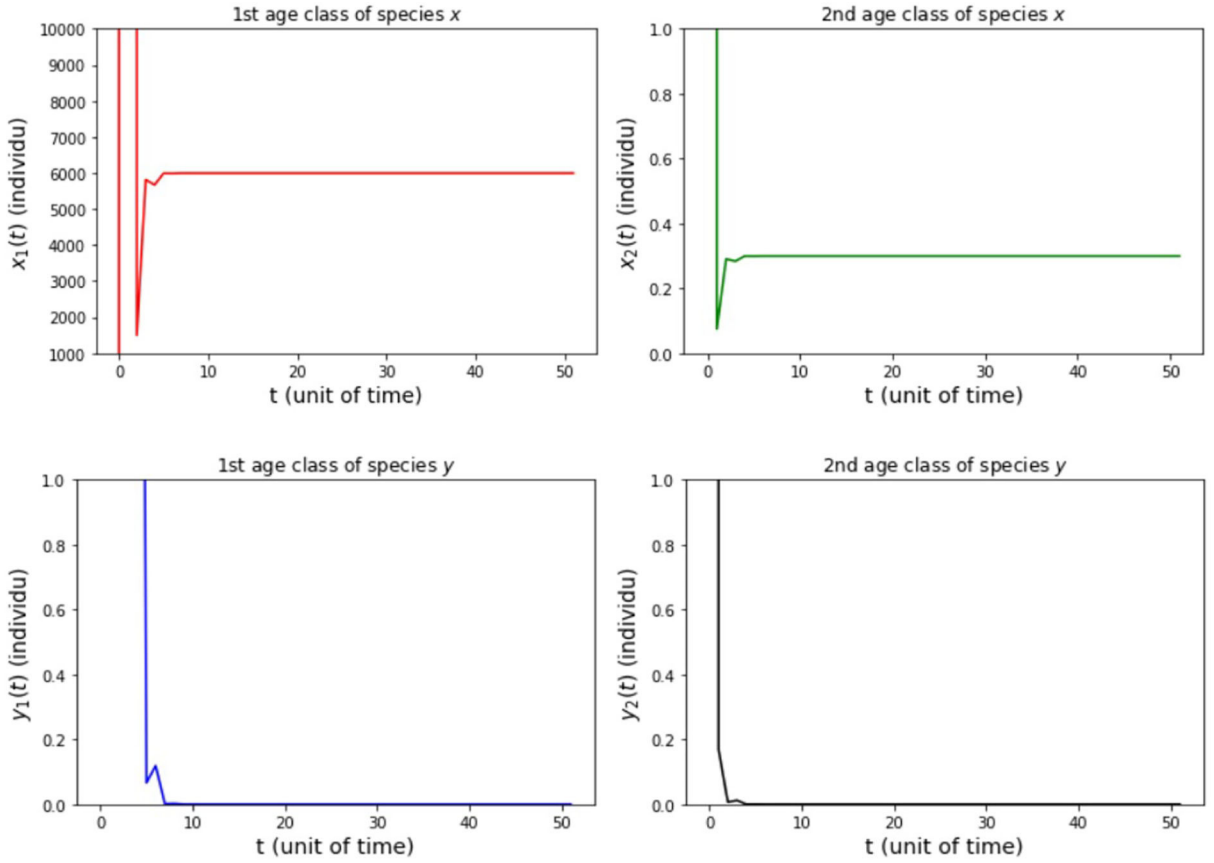


FIGURE 3
Population growth of each age class from case (ii) Model A.

results obtained

$$
\begin{aligned}
|g_{11}| &= 1 > 0,\\
\begin{vmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{vmatrix} &= -\frac{a(a(1-R_x)-b(1-R_y))}{R_x(a^2-b^2)},\\
\begin{vmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{vmatrix} &= -\frac{a(a(1-R_x)-b(1-R_y))}{R_x(a^2-b^2)},\\
|G| &= \frac{(a(1-R_y)-b(1-R_x))(a(1-R_x)-b(1-R_y))}{R_xR_y(a^2-b^2)}.
\end{aligned}
$$

The results above show that only $|g_{11}|$ is positive, so there are three minor principals that must be determined to be positive. The three minor principals will be positive if $(a^2 - b^2) > 0, a(-R_x + 1) - b(-R_y + 1) < 0$, and $a(-R_y + 1) - b(-R_x + 1) < 0$. Thus, the equilibrium point E_{xy} is locally stable asymptotically if $a(a^2 - b^2) > 0, a(-R_x + 1) - b(-R_y + 1) < 0$, and $a(-R_y + 1) - b(-R_x + 1) < 0$. The proof is complete. ■

4. Numerical simulations

In the previous sections, analytical analyses of the existence and locally stable asymptotically of each equilibrium point have been carried out in Models A and B. This section presents some numerical simulation results with the aim of providing numerical proof of Theorems 1 and 2 of this study. The graphical results of the numerical proof are presented in Figures 2–8. In this study, the specific species in question is not determined. For this reason, some of the parameter values used in the numerical simulation are hypothetical parameters. In this case, simulations are performed on both models and each model will be divided into several cases based on the stability conditions of each equilibrium point.

In the numerical simulation of Model A, the simulation is divided into three cases. It is assumed for all the cases that the

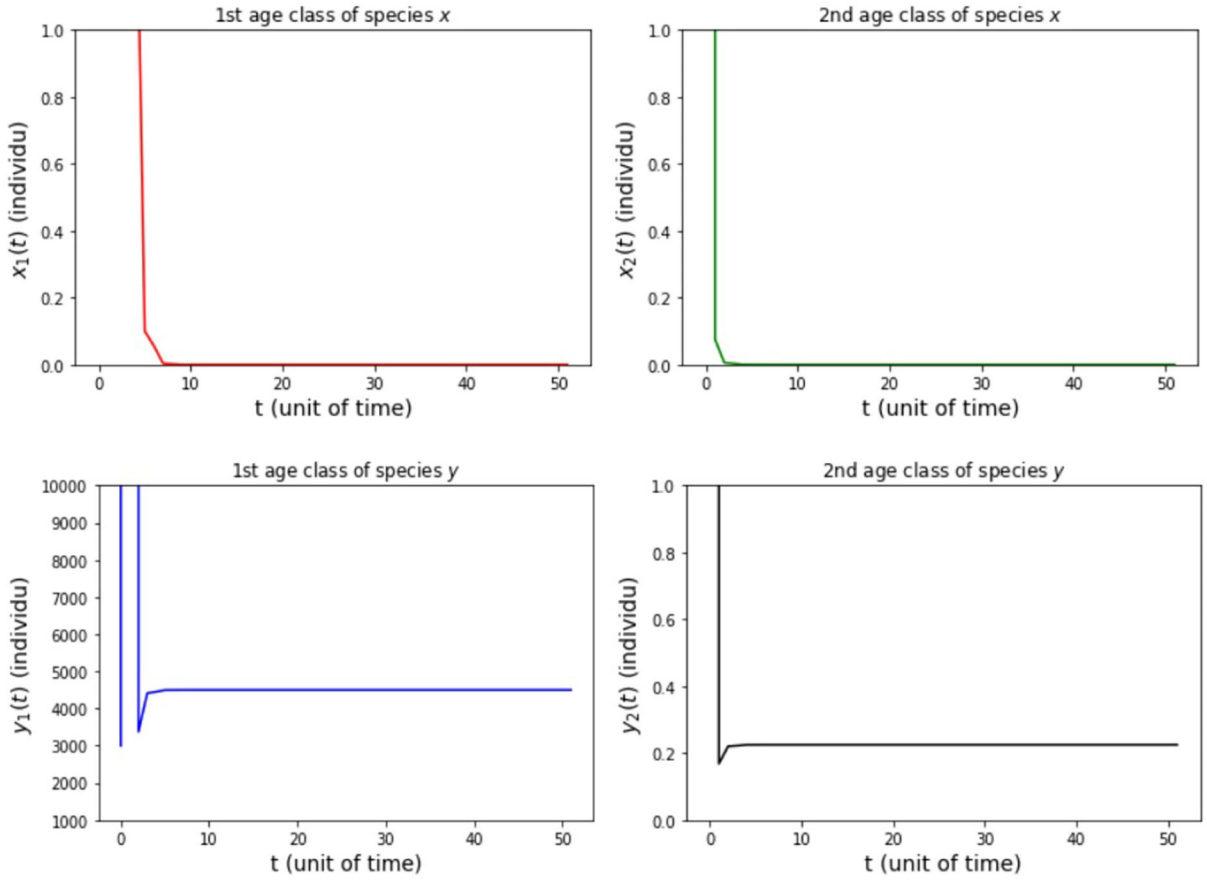


FIGURE 4
Population growth of each age class from case (iii) Model A.

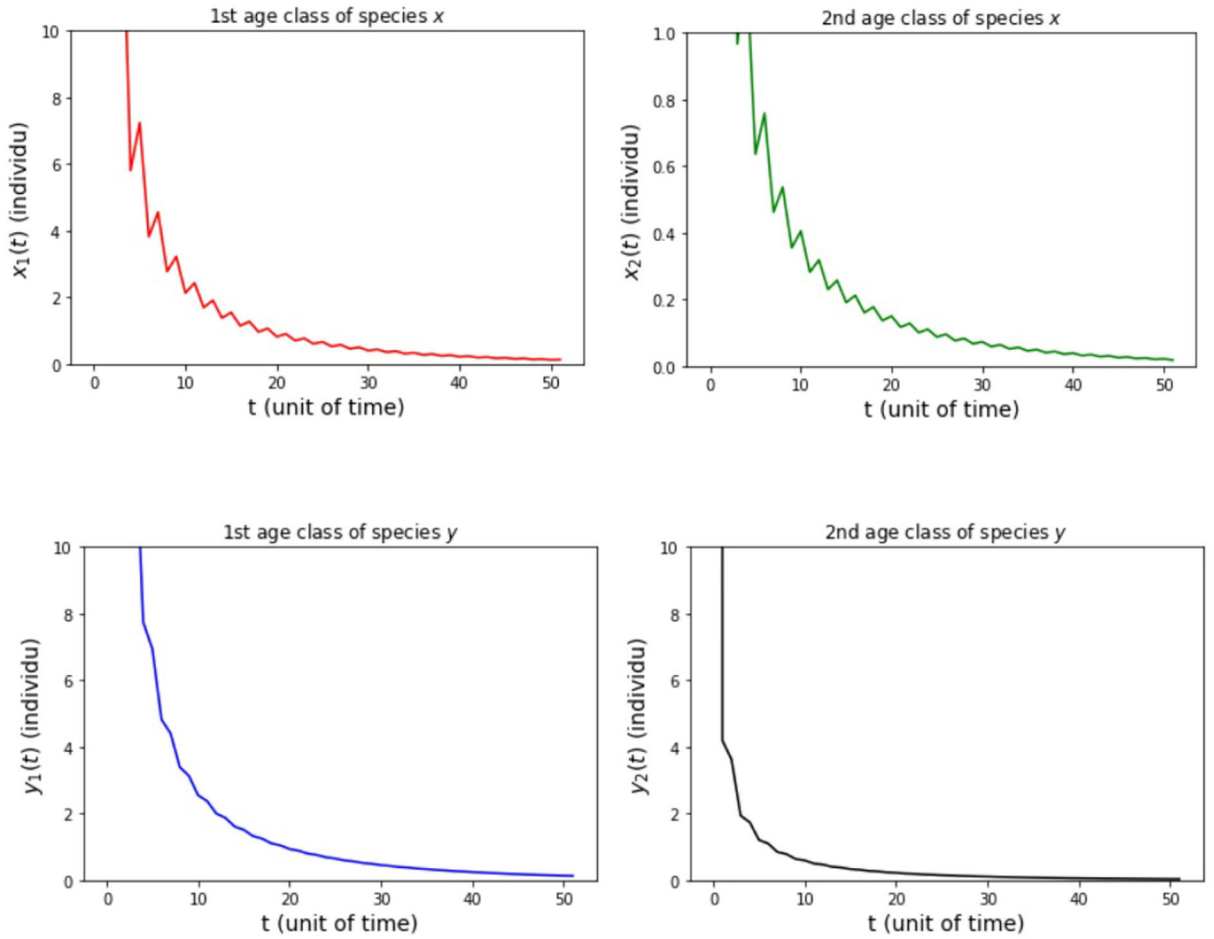


FIGURE 5
Population growth of each age class from case (i) Model B.

parameter values are $s_{x_1} = 0.3$ and $s_{y_1} = 0.3$. Both of these parameters were obtained in the research of Travis et al. [7]. The other parameters for each case in the numerical simulation of Model A 275 is presented as follows:

(i) $f_{x_2} = 6, f_{y_2} = 4, h_{x_2} = 1/2$, and $h_{y_2} = 1/4$ so that $R_x = 0.89$ and $R_y = 0.89$.
(ii) $f_{x_2} = 20,000, f_{y_2} = 500, h_{x_2} = 0.001$, and $h_{y_2} = 1/4$ so that $R_x = 5,994$ and $R_y = 22.5$.
(iii) $f_{x_2} = 500, f_{y_2} = 20,000, h_{x_2} = 0.001$, and $h_{y_2} = 1/4$ so that $R_x = 149.85$ and $R_y = 4500$.

The parameter value for the birth rate of 20,000 follows the value of the birth rate parameter in the study of Travis et al. [7]. All parameter values are measured per unit of time and the total population is calculated per individual.

The results of simulation on Model A for cases (i)-(iii) are presented in Figures 2–4. Figure 2 shows the case (i), in which $R_x < 1$ and $R_y < 1$, resulting in the stability of the system toward the equilibrium point E_0. Figure 3 shows case (ii), in which $R_x > 1$ and $R_x > R_y$, resulting in the stability of the system toward the equilibrium point $E_x = [5993, 0.3, 0, 0]^T$, i.e., species x exists. Figure 4 shows case (iii), in which $R_y > 1$ and $R_y > R_x$, resulting in the stability of the system toward the equilibrium point $E_y = [0, 0, 4499, 0.22]^T$ i.e., species y exists.

In the numerical simulation of Model B, the simulation is divided into four cases. It is assumed that the survival rate of Model B is the same as the survival rate of Model A. It is also assumed that the level of intraspecific competition is $a = 0.05$ and the level of interspecific competition is $b = 0.01$. The values of a and b indicate that $a^2 - b^2 > 0$. This choice was made to minimize the simulations carried out. Then, we assume the value of the parameters $h_{x_2} = 1/2$ and $h_{y_2} = 1/4$ for all cases of Model B. The other parameters for each case in the numerical simulation of Model B are presented as follows:

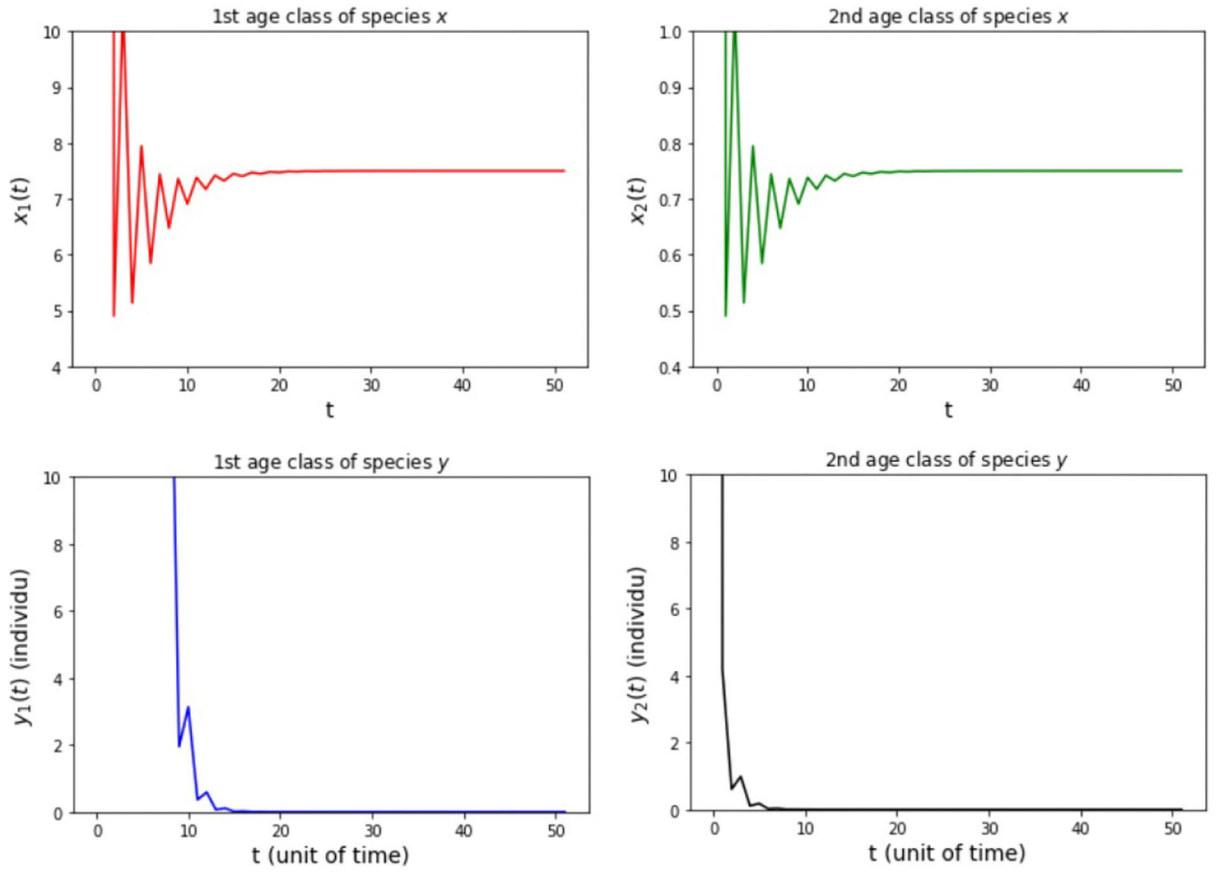


FIGURE 6
Population growth of each age class from case (ii) Model B.

(i) $f_{x_2} = 6$ and $f_{y_2} = 4$ so that $R_x = 0.89$ and $R_y = 0.89$.
(ii) $f_{x_2} = 20{,}000$ and $f_{y_2} = 500$ so that $R_x = 3{,}000$ and $R_y = 112.5$.
(iii) $f_{x_2} = 500$ and $f_{y_2} = 20{,}000$ so that $R_x = 75$ and $R_y = 4{,}500$.
(iv) $f_{x_2} = 20{,}000$ and $f_{y_2} = 10{,}000$ so that $R_x = 3{,}000$ and $R_y = 2{,}250$.

The results of the numerical simulation on Model B are shown in Figures 5–8. Figure 5 shows case (i), in which $R_x < 1$ and $R_y < 1$, resulting in the stability of the system toward the equilibrium point E_0. Figure 6 shows the case (ii), in which $R_x > 1$ and $a(1 - R_y) > b(1 - R_x)$, resulting in the stability of the system toward the equilibrium point $E_x = [59980, 3, 0, 0]^T$ where species x exists. Figure 7 shows case (iii), in which $R_y > 1$ and $a(1 - R_x) > b(1 - R_y)$, resulting in the stability of the system toward the equilibrium point $E_y = [0, 0, 89980, 4.495]^T$. Figure 8 shows case (iv), in which $a > b$, $a(1 - R_x) < b(1 - R_y)$, and $a(1 - R_y) < b(1 - R_x)$, resulting in the stability of the system toward the equilibrium point $E_{xy} = [53108.33, 2.65, 34358.33, 3.43]^T$.

5. Discussion

The research presented in this paper is an extension of the model carried out by Pennycuick et al. [6], Travis et al. [7], and Kon [8–10]. In these studies, the effect of harvesting on species growth which is influenced by density-dependence and competition has not been studied. These influences on the growth of species can occur in an ecosystem as in the case study on fish conducted by Travis et al. [7] in his research. This extended model can also be applied to the species growth model without the effect of harvesting with the value of the harvesting level equal to zero.

In this paper, we find several results in the comparison of the models we developed, namely Model A and Model B. In Model A, there is no co-existence equilibrium point between the two semelparous species in the same ecosystem, but this point has only been found in Model B. It can be concluded that the competition affects the existence of a co-existence equilibrium point in the density-dependent growth of the two species. In addition, the local stability in Model A and Model B is also

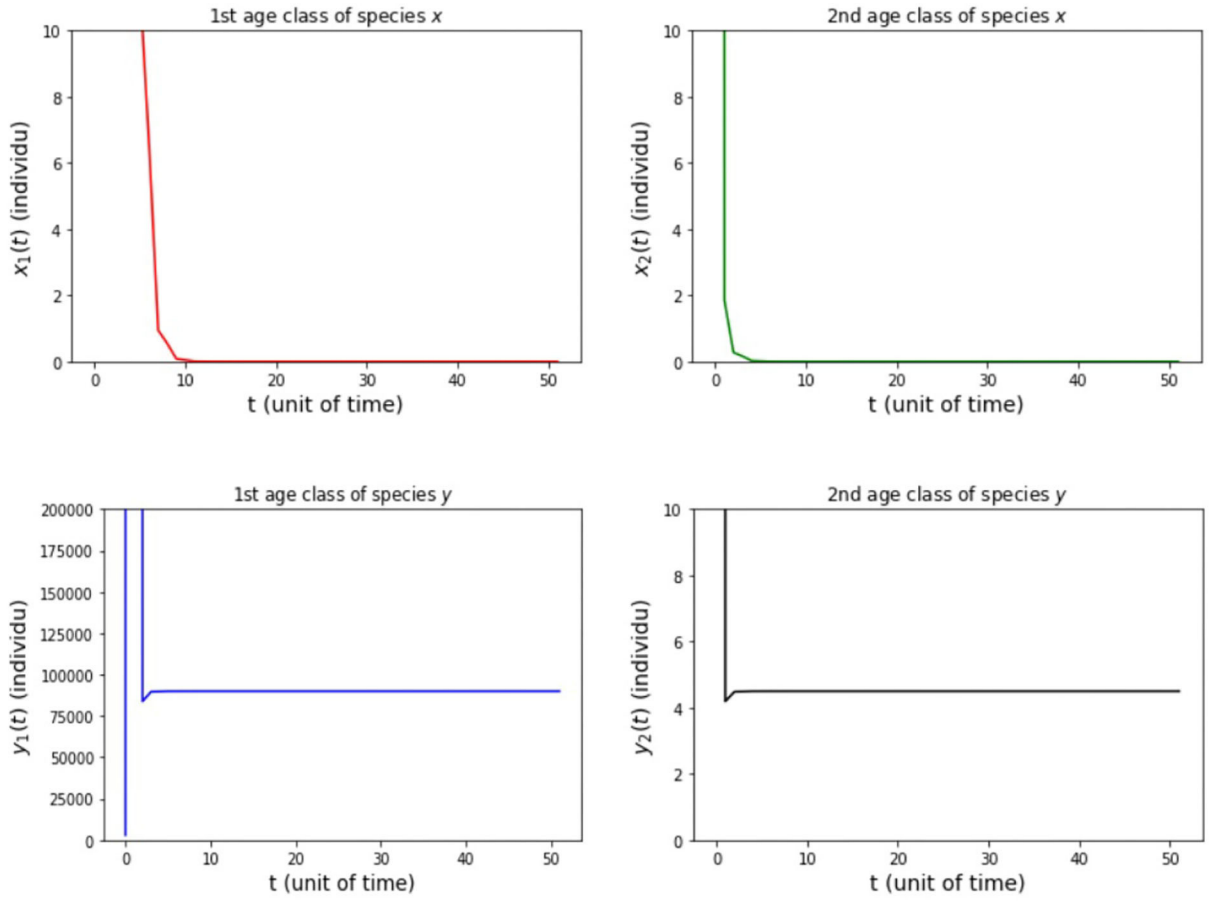


FIGURE 7
Population growth of each age class from case (iii) Model B.

explored. The asymptotically local stability of Model A depends on the condition of the inherent net reproduction number of species x (R_x) and the inherent net reproduction number of species y (R_y). Then, the asymptotically local stability of Model B depends on the values of R_x, R_y, the level of intraspecific competition (a), and the level of interspecific competition (b).

6. Conclusion

In this paper, population growth models of two semelparous species have been studied where growth is influenced by density-dependent and harvesting. The models are analyzed with no competition as in Model A and with competition as in Model B. Based on the results, the factors that influence both the existence of the equilibrium point and its locally stable asymptotically are the values of a, b, R_x (net reproductive value of species x with harvesting effect), and R_y (net reproductive value of species y with harvesting effect). Overall, both the equilibrium points and their local stability are determined by the parameters R_x and R_y, which implies that harvesting is very influential. Besides, in Model A, there is no equilibrium point where both species exist. Therefore, Model A was developed into Model B and it turned out that adding the factors of intraspecific competition (a) and interspecific competition (b) affect the equilibrium point where the two species require that $a \neq b$. Then, the values of parameters a and b also play a role in determining the local stability conditions from the equilibrium point where both species exist. Observations on multispecies growth are closer to the real problem than single species growth. However, the multispecies growth model is difficult to work with. Hence, there is a simplification of the particular model currently being investigated. This study still focuses on the simple problem of the growth of two semelparous species that are affected by density-dependent, competition, and harvesting. Future work that can be developed from this research include research on the number of any age class, the number of any species, global stability problems, bifurcation problems, and many more.

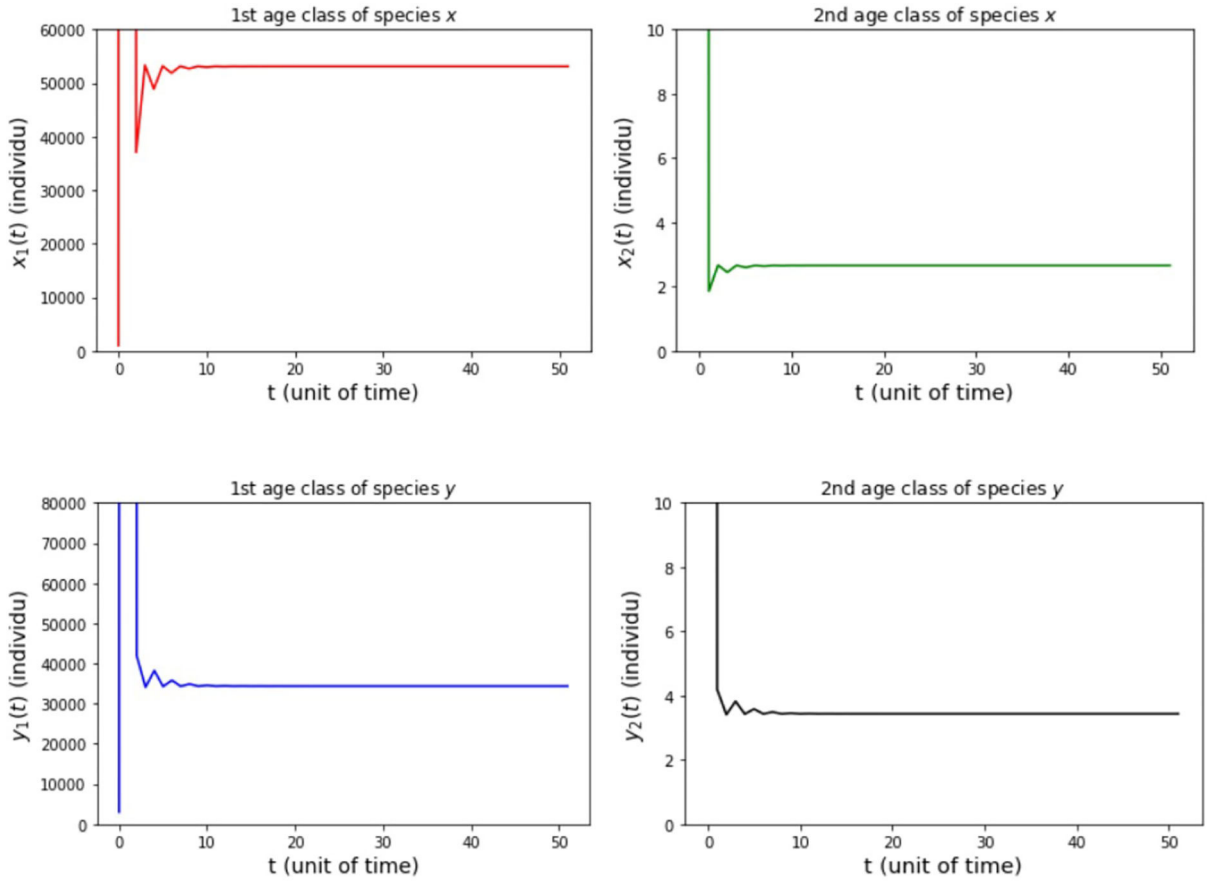


FIGURE 8
Population growth of each age class from case (iv) Model B.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

Author contributions

AS: design and supervise the research. AH: develop and analyze the models. EC: conducting literature review. All authors contributed to the article and approved the submitted version.

Funding

The authors would like to thank the Indonesia Government who provided fund for the APC of the publication in this journal through the scheme of Penelitian Tesis Magister (PTM)-2022 with contract number 1318/UN6.3.1/PT.00/2022.

Acknowledgments

The work is part of the research during the M.Sc. candidature of AH. AH would like to thank for the tuition fee during the candidature covered by ALG-Unpad Research Grant for the year 2020-2021 with contract number 1959/UN6.3.1/PT.00/2021.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Chow Y, Kon R. Global dynamics of a special class of nonlinear semelparous Leslie matrix models. *J Diff Equat Appl*. (2020) 26:1777288. doi: 10.1080/10236198.2020.1777288

2. Diekmann O, Planqué R. The winner takes it all: how semelparous insects can become periodical. *J Math Biol*. (2020) 80:283–301. doi: 10.1007/s00285-019-01362-3

3. Diekmann O, Davydova N, Van Gils S. On a boom and bust year class cycle. *J Diff Equat Appl*. (2005) 11:327–35. doi: 10.1080/10236190412331335409

4. Davydova NV, Diekmann O, Gils SAV. Year class coexistence or competitive exclusion for strict biennials? *J Math Biol*. (2003) 46:95–131. doi: 10.1007/s00285-002-0167-5

5. Leslie PH. On the use of matrices in certain population mathematics. *Biometrika*. (1945) 33:183. doi: 10.1093/biomet/33.3.183

6. Pennycuick CJ, Compton RM, Beckingham L. A computer model for simulating the growth of a population, or of two interacting populations. *J Theor Biol*. (1968) 18:316–29. doi: 10.1016/0022-5193(68)90081-7

7. Travis CC, Post WM, DeAngelis DL, Perkowski J. Analysis of compensatory Leslie matrix models for competing species. *Theor Popul Biol*. (1980) 18:16–30. doi: 10.1016/0040-5809(80)90037-4

8. Kon R. Age-structured Lotka-Volterra equations for multiple semelparous populations. *SIAM J Appl Math*. (2011) 21:694–713. doi: 10.1137/100794262

9. Kon R. Permanence induced by life-cycle resonances: the periodical cicada problem. *J Biol Dyn*. (2012) 6:855–90. doi: 10.1080/17513758.2011.594098

10. Kon R. Stable bifurcations in multi-species semelparous population models. In: Elaydi SMHHY, Potzsche C, editor. *Springer Proceedings in Mathematics and Statistics. Vol. 212*. New York, NY: Springer New York LLC (2017). p. 3–25.

11. Beay LK, Suryanto A, Suryanto A, Suryanto A. Stability of a stage-structure Rosenzweig-MacArthur model incorporating Holling type-II functional response. In: *IOP Conference Series: Materials Science and Engineering*. (2019). Available online at: https://iopscience.iop.org/article/10.1088/1757-899X/546/5/052017

12. Savadogo A, Sangaré B, Ouedraogo H. A mathematical analysis of Hopf-bifurcation in a prey-predator model with nonlinear functional response. *Adv Diff Equat*. (2021) 2021:275. doi: 10.1186/s13662-021-03437-2

13. Fang L, N'gbo N, Xia Y. Almost periodic solutions of a discrete Lotka-Volterra model via exponential dichotomy theory. *Am Inst Math Sci*. (2022) 7:3788–801. doi: 10.3934/math.2022210

14. Wikan A. Dynamical consequences of harvest in discrete age-structured population models. *J Math Biol*. (2004) 49:35–55. doi: 10.1007/s00285-003-0251-5

15. Cooke KL, Elderkin R, Witten M. Harvesting procedures with management policy in iterative density-dependent population models. *Nat Resour Model*. (1988) 2:383–420. doi: 10.1111/j.1939-7445.1988.tb00065.x

16. Haight RG, Getz WM. *Population Harvesting: Demographic Models of Fish, Forest, and Animal Resources*. Princeton, NJ: Princeton University Press (1989).

17. Ganguli C, Kar TK, Mondal PK. Optimal harvesting of a prey-predator model with variable carrying capacity. *Int J Biomath*. (2017) 10:17500693. doi: 10.1142/S1793524517500693

18. Pratama RA, Ruslau MFV, Nurhayati. global analysis of stage structure two predators two prey systems under harvesting effect for mature predators. In: *Journal of Physics: Conference Series. Vol. 1899*. IOP Publishing Ltd (2021). Available online at: https://iopscience.iop.org/article/10.1088/1757-899X/546/5/052017

19. Hannesson R, Flåten O, Hannesson R, Flaten O. The economics of multispecies harvesting. *Scand J Econ*. (1989) 91:340091. doi: 10.2307/3440091

20. Supriatna AK, Possingham HP. Harvesting a two-patch predator-prey metapopulation. *Nat Resour Model*. (1999) 12:481–98. doi: 10.1111/j.1939-7445.1999.tb00023.x

21. Supriatna AK. Maximum sustainable yield for marine metapopulation governed by coupled generalised logistic equations. *J Sustain Sci Manag*. (2012) 7:201–6. Available online at: https://jssm.umt.edu.my/files/2012/11/13.pdf

22. Husniah H, Supriatna AK. Marine biological metapopulation with coupled logistic growth functions: the MSY and quasi MSY. In: *AIP Conference Proceedings*. Makassar (2014). Available online at: https://aip.scitation.org/doi/10.1063/1.4866532

23. Supriatna AK, Husniah H. Sustainable harvesting strategy for natural resources having a coupled Gompertz production function. In: *Interdisciplinary Behavior and Social Sciences - Proceedings of the 3rd International Congress on Interdisciplinary Behavior and Social Sciences, ICIBSoS 2014*. Bali (2015). Available online at: https://www.taylorfrancis.com/chapters/edit/10.1201/b18146-19/sustainable-harvesting-strategy-natural-resources-coupled-gompertz-production-function-supriatna-husniah

24. Supriatna AK, Ramadhan AP, Husniah H. A decision support system for estimating growth parameters of commercial fish stock in fisheries industries. *Procedia Comput Sci*. (2015) 59:331–9. doi: 10.1016/j.procs.2015.07.575

25. Husniah H, Anggriani N, Supriatna AK. System dynamics approach in managing complex biological resources. *ARPN J Eng Appl Sci*. (2015) 10. Available online at: https://www.arpnjournals.com/jeas/research_papers/rp_2015/jeas_0315_1650.pdf

frontiers | Frontiers in Applied Mathematics and Statistics

TYPE Review
PUBLISHED 01 December 2022
DOI 10.3389/fams.2022.1020353

OPEN ACCESS

EDITED BY
Mohd Hafiz Mohd,
Universiti Sains Malaysia (USM),
Malaysia

REVIEWED BY
Christophe Le Page,
UMR SENS-CIRAD, France
Hirohide Haga,
Doshisha University, Japan

*CORRESPONDENCE
Arthur Brugière
arthur.brugiere@ird.fr

SPECIALTY SECTION
This article was submitted to Mathematics of Computation and Data Science, a section of the journal Frontiers in Applied Mathematics and Statistics

RECEIVED 16 August 2022
ACCEPTED 04 November 2022
PUBLISHED 01 December 2022

CITATION
Brugière A, Nguyen-Ngoc D and Drogoul A (2022) Handling multiple levels in agent-based models of complex socio-environmental systems: A comprehensive review. *Front. Appl. Math. Stat.* 8:1020353. doi: 10.3389/fams.2022.1020353

Handling multiple levels in agent-based models of complex socio-environmental systems: A comprehensive review

Arthur Brugière[1,2]*, Doanh Nguyen-Ngoc[1,2] and Alexis Drogoul[1,2]

[1]UMI 209 UMMISCO, IRD, Sorbonne Université, Bondy, France, [2]International Joint Laboratory ACROSS, IRD, Thuyloi University, Hanoi, Vietnam

Agent-based modeling (ABM) has been successfully used, since its emergence in the 1990s, to model and simulate the dynamics at work in complex socio-environmental systems, in many domains and applications where interactions between people and their environments give rise to emergent phenomena that are difficult to study otherwise (urban planning, land-use change, adaptation to environmental changes, biodiversity protection in socio-ecosystems, environmental pollution control, etc.). The inclusion of multiple levels of analysis, abstraction, and representation in these models, however, is much more recent and is still the subject of many proposals and discussions within a relatively informal field, Multilevel Agent-Based Modeling (ML-ABM), which is most often presented as an approach that extends the classical ABM paradigm to include multilevel concepts. Over the past decade, ML-ABM has been increasingly adopted and explored by researchers as an effective paradigm for framing and defining the mechanisms underlying multilevel dynamics. However, due to the youth of the field, no single definition, methodology, or tool unifies studies in this rapidly expanding area. This review will begin with an introduction to socio-environmental systems (SES) and the challenges that modeling approaches face in representing them properly, especially regarding the complexity of human behaviors and organizations. ABM presents opportunities for modeling SESs with respect to these challenges, including the simulation of individual and social behavior and their ability to provide a descriptive and generative representation of the simulated system. However, ABM is limited in its ability to represent levels and scales, as these concepts are absent from the classical ABM metamodel. A complete review of the ML-ABM literature will be carried out, structured around a continuum that emerged during the review: that of the distribution of behaviors (and thus, from a software engineering perspective, of control) across the levels, from approaches that allow only one level to be active at a time, to approaches that rely on simultaneous activity and feedback loops between several levels. Different design choices will, thus, be presented to meet the different needs of multi-level representation, focusing on the interest on modelers and the strengths and limitations of each. In particular, we will highlight a limitation shared by all the reviewed approaches, namely their inability to represent several parallel hierarchies of levels and their interactions, a capability that

appears more and more crucial to finely represent social behaviors in SES. A new perspective on the interest that the AGR approach could represent to allow this representation of hierarchies allows us to conclude on the research perspectives are still open.

KEYWORDS

review, multi-level agent based model, design pattern, socio-environmental systems (SES), multi-level, agent-based modeling

1. Introduction

1.1. Socio-environmental systems

We have been living in the so-called Anthropocene [1] for some time now, where human influence on the environment is expanding globally and driving the need for holistic and integrated approaches to understanding coupled natural and human systems, which includes monitoring, analyzing, and modeling the complex systems resulting from their interactions.

This is where the concept of a social-environmental system (or socio-ecological system), SES, comes in. It is defined as a complex dynamic system that includes people and nature and is continually changing in response to internal or external pressures [2]. Pressures can result from the behavior of their social (e.g., demographic changes) and ecological (e.g., climate fluctuations) components, or, more usually, from a combination of both (e.g., human-induced climate change).

However, as interesting as this concept is for describing human-nature complex systems, it significantly defies available modeling capabilities. On one hand, SES is not really different from "ordinary complex systems," i.e., systems whose perceived complicated behaviors can be attributed to one or more of the following characteristics: a large number of possibly heterogeneous elements, numerous possibly non-linear and discontinuous relationships between these elements, a dynamic emergence of forms (phenomena, structures, aggregates, organisms, or problems), at different levels of abstraction or scales, to which the elements adapt [3].

This notion of "emergence," and the related notions of "levels" or "scales" lie at the heart of the various complexity theories [4], where the spontaneous appearance of structured macroscopic patterns resulting from independent microscopic interactions between the elements of the system is considered as a distinctive signature that distinguishes simple systems from complex systems.

The "scales" used in complex systems usually refer to three possible quantifiable dimensions: space, time, and size [5], which are then used to define and position "levels" at which the system can be observed, described, or analyzed. It is common to distinguish between micro, meso, and macro levels, using the hierarchical view of systems and subsystems popularized by Herbert Simon in his famous article "The Architecture of Complexity" [6]. According to this vision, a large number of indivisible individual elements, engaged in local short-term interactions, compose the micro-level; as one progresses to the higher levels, the spatial and temporal scales become larger, and the number of elements smaller. Each level, except the macro-level, is nested into or spatially and temporally bounded by a higher level.

These levels may be explicit in the description of the system because they have a recognized existence in reality (e.g., an organism, a physical structure) or because they correspond to scales at which observers want to describe or understand the system. The passage between these levels, similarly, may be described explicitly in the description of the system (e.g., in the form of a count or aggregation or disaggregation functions between levels), or it may be considered emergent. Here, "emergence" refers to the way in which structural or functional properties can emerge spontaneously, in an unplanned manner, at a given level from the self-organization of elements identified at a finer level [7], this notion of self-organization refers to a bottom-up process in which a system modifies its internal organization to adapt to changes in its goals and the environment without explicit external or top-down control.

All these characteristics translate into a real difficulty in producing an analytical and deterministic description of the behaviors of complex systems, as models of these systems imply being able to represent non-linear, ill-posed, or chaotic behaviors operating simultaneously at multiple levels of abstraction. But there are also specific characteristics of SES, mainly related to the fact that they include human elements, which make this category of systems much more complicated to deal with using traditional modeling techniques.

1.2. Challenges of SES modeling

Socio-environmental systems modeling (SES) consists of developing models to study the complex problems that arise from interactions between human (i.e., social, economic)

and natural (i.e., biophysical, ecological, and environmental) systems. While there are many ways to describe SEA (especially in the anthropological or sociological literature), dynamic models appear to be indispensable tools for understanding and managing them, as they allow us to study the behavior of these systems not only under past and present conditions but also in the future, through scenario analysis or the virtual exploration of possible paths. With the advent of new techniques and computational power on one hand and the growing sustainability challenges on the other, SES modeling is used to support multiple goals, such as informing decision-making and science or raising awareness, or promoting education and communication on sensible issues [8]. However, to be useful, these models must be able to handle, in an integrated way, the complexity of SES, often characterized by intricate feedback loops in human-nature and human-society interactions [9, 10], at different levels of abstraction and with different levels of details [11]. More generally, they need to address the eight major SES modeling challenges listed by Elsawah et al. [8]:

1. Bridging epistemologies across disciplines.
2. Integrated treatment of modeling uncertainty.
3. Combining qualitative and quantitative methods and data sources.
4. Dealing with scales and scaling.
5. Capturing Systemic changes in SES.
6. Integrating the human dimension.
7. Elevating the adoption of SES models.
8. Leveraging new data types and sources.

Reading the article and the arguments behind each of the "challenges," it is easy to see that the first three, which concern interdisciplinary, uncertainty, and source diversity, are not at all unique to SES modeling, but are, of course, shared by any type of integrated modeling involving different domains. Similarly, the last two, which are concerned with the adoption and integration of models in decision-making, are largely shared by the whole field of modeling, whether integrated or not. This leaves challenges 4, 5, and 6 as specific to the modeling of SES, i.e., in short, those related to the representation of human behavior (individual and social) and to the representation of different levels of abstraction between which emergencies and "systemic" constraints can be identified.

Models of SES have been developed using diverse approaches, including system dynamics, Bayesian networks, agent-based models, dynamic stochastic equilibrium models, statistical micro-simulation models, and any hybridization of these methods [12], but only one so far, agent-based modeling (ABM), has proved its capacity to faithfully represent individual and social human behaviors.

1.3. ABM and SES modeling

Agent-based modeling effectively present clear opportunities for SES modeling and help to address some of the main weaknesses of the various categories of models described in Schulze et al. [13] and Giupponi et al. [2], particularly with respect to the simulation of individual and social behavior and their ability to provide a descriptive and generative representation of the simulated system along the four dimensions briefly described below (loosely based on Giupponi et al. [2]).

The first is heterogeneity

In general, ABMs consist of detailed dynamic simulations in which many heterogeneous human and natural agents interact: it avoids a coarse, average, and, therefore, unrealistic representation of the system components. Human agents may vary in their demographic characteristics, location, endowments, individual capabilities, worldview, attitudes, and behavior. Natural agents can also vary in spatial and temporal attributes.

The second is individual complexity

Compared to natural agents, human agents are more complex to simulate, as they perform deliberative processes and make autonomous individual decisions [14, 15]. Behavioral complexity arises from agents' mental models [16, 17] or more commonly their "architectures," which include their cognition, reasoning, and learning capabilities, based on the abundance of social science theories about how human agents behave in various contexts. ABM has the potential to enable the exploration of this set of decision-making theories, including the ability of agents to learn from past experiences [18], which is extremely important for long-term simulations of SES evolution.

The third is Interactions and, in particular, social interactions

Not only are human agents deliberative, but they are also social: they communicate with other agents and their behavior arises from interactions in multiple contexts with other human and nonhuman agents and the environment [19]. This aspect is fundamental capturing dynamics such as clustering, imitation, learning, and diffusion processes. It is a crucial feature for modeling SES insofar as the interactions of agents, and in particular informal relationships and opinion dynamics, can shape collective patterns of behavior.

The fourth dimension deals with the representation of organizations and the emergence of organizational structures

Human agents are deliberative and social, but they are also organizational and can form themselves into various structures, hierarchical or not. At the same time, norms and institutions, whether fixed or emergent, may induce individuals to act differently from their individual choices, which may be crucial to understanding the appearance or disappearance of certain dynamics.

The potential of ABM is thus clear for representing the social, heterogeneity, and interaction dimensions of SES, and indeed is the approach of choice for the majority of researchers faced with this need [2], but considerable methodological challenges remain, in particular, as will be seen in this review, those related to the explicit representation of scales and levels of abstraction, absent from the classical meta-model of agent-based modeling [20].

1.4. Presentation of the review

This review will explore how multi-level agent-based modeling (ML-ABM) has evolved and what solution it proposes to address modelers' representation needs. The aforementioned propositions (either formalism, framework, or *ad hoc* implementations) will be organized along a continuum expressing the autonomy of the levels in terms of control.

Along this continuum will be described three milestones of interest. At one end of the continuum, this review starts with the pattern we name Section 2.2, the most restrictive in terms of the autonomy of levels, as it can only represent one active level at a time. It is followed, in the middle, by a pattern we name Section 2.3, which allows the levels to have a certain degree of autonomy in terms of behavior, but imposes to describe strict hierarchical coordination over these behaviors; then, at the last end of the continuum, the one we name Section 2.4 giving levels a high degree of autonomy and control, in exchange of some complexification in the description of their coordination.

In the course of the review, the strengths and limitations of these approaches will be highlighted in relation to the needs expressed by modelers of complex socio-environmental systems in terms of modeling and simulation, as well as from a software engineering perspective. Specifically, in terms of modeling, the review focuses on how each pattern supports the representation of levels and their interactions or feedback, the design of the overall architecture, and the ease of use for modelers. In terms of simulation, it focuses on the operational instantiation of processes such as the consideration of emergence phenomena between levels.

This review will initiate a discussion of the common limitations of the patterns. It will first point out that ABM lacks an explicit representation of time and space scales, which are necessary for ML-ABM. It will then outline the difficulty of escaping the "single viewpoint hierarchy" to describe and simulate phenomena where multiple hierarchies (social, environmental, etc.) may be involved at the same time.

This discussion will allow us to introduce the proposal of a more faithful representation of multiple scales in ABM (Section 3.2). Although ABM does not really provide a way to represent the entirety of the concepts used in complex systems science (notably simultaneous hierarchies of viewpoints), its software basis, multi-agent systems (MAS), had explored some interesting paradigms that could be used to enrich the patterns presented in this review. This will be illustrated, among others, with the AGR architecture and its main change being the use of roles to describe groups/levels and the aggregation of roles to describe agents. This would remove the obligation to have agents belonging to only one super-level but would also create more complexity for modelers to use and analyze emerging patterns.

Finally, it will conclude by summarizing the studies presented in the papers in relation to the necessities expressed by modelers trying to build models of complex socio-environmental systems. Then it lists their respective limitations and calls for renewed research in this area using the approach presented in the discussion.

2. Multi-level ABM in the literature

2.1. Introduction

We have seen in Section 1.3 that ABM has the potential to address some of the challenges raised by the modeling of SES, particularly the three points listed below:

4. Dealing with scales and scaling.
5. Capturing Systemic changes in SES.
6. Integrating the human dimension.

The way they are addressed in the ABM literature is of course highly dependent on the needs of modelers and the goals of the models themselves; not all the modeling questions require to represent simultaneously multiple levels or the emergence of structures and functions, which is why there exist different architectures for implementing "multi-level" ABMs. This multiplicity of offers is the reason why the vocabulary is not completely fixed, making it sometimes difficult to find direct correspondences between the concepts used in complex systems science and the ones proposed in the different ML-ABM approaches. For instance, instead of using concepts of time and space scales, ABM designers have to deal with scheduling algorithms and encapsulation to translate temporal and spatial constraints and transfers between agents that represent levels; similarly, "systemic changes" can be represented by different aspects: dynamic instantiation of agents, injection of new code, etc.

Implementing these levels and their interactions to represent scales and their links is then dependent on two aspects: on one hand, the choices of the modelers regarding the resulting complexity of the models, which ultimately depends on their usage: should they be kept simple enough to be understood by anyone? Should they exhibit emerging properties only found in "real" complex systems? Should they allow different levels of exploration and explanation?; on the other hand, of course, the technical and computational limitations of the existing languages and computer systems available for implementing the ABM architectures used. Both aspects have evolved over time, and although we can witness a complexification of the questions asked to models in parallel with a considerable increase in computational capabilities devoted to running simulations, it does not mean that "simple" approaches are not relevant in most cases, which explains the diversity of the proposals listed in this review. Morvan and Jolly [21] reviewed those points and demonstrate that it has been applied in numerous different research fields as biomedical research, human flow, social science, ecology, etc.

Some previous study did draw some categories of ML-ABM trying to give a reading grid over previous modeling questions. Notably, Mathieu et al. [22] extracted four categories based on modelers' coupling choices (following a decision tree) between two levels. This approach represents a valuable tool for modelers building ML-ABM but does not give more detail on the overall internal model mechanism. This review provides a complementary analysis and aims to see existing ML-ABM architectures, frameworks, and approaches from the point of view of level control in the global model.

As displayed in Figure 1, we propose to present the evolution of level control along a continuum. This continuum goes from completely rigid (Zoom) without any autonomy of the levels to levels not completely autonomous but running under rigid control (Russian Dolls) or autonomous levels with relatively flexible control (Collaboration). Those three milestones are not impermeable, and there are some models and work in between them.

These levels' control differences can come from the conceptualization, creation, and organization of levels in a multi-level architecture, which is handled differently between the different approaches presented here. The Section 2.2 design pattern which coordinates levels as different models, each built with a single temporal and spatial scale, uses functions to allow switching from one level to another. Some other approaches aim to have a more discrete composition between levels and include each level into another one in a spatio-temporal hierarchical order. This extension can keep ABM's scheduling concept, like with the Section 2.3 pattern which allows building an integrated model at the cost of losing the level's inner control, or, without this control lost, as in the Section 2.4 pattern weak coupling model's level.

All these level control difference creates some extra complexity in the exploitation of those models (point 5 from Elsawah et al. [8]). When a model starts to have different levels (made or not of different independent models), it becomes more complicated to capture and analyze emerging patterns from the global system or in-between described levels. This analytic point can be needed by modelers which require to keep detailed information from each level and ensure interaction between them. Russian Dolls and Collaboration patterns provide different approaches to do so. The first one uses similar analysis tools as classical ABM as this multi-level architecture remains very close to it. The second one uses some tools inspired by software engineering.

But, some multi-level models do not need this to analyze and capture those changes, as in the first multi-level model presented by Gil-Quijan et al. [5] about the growth of a cancerous tumor. At the beginning of the simulation, the micro-level (at the cell scale) is important as it lets emerging a cluster of cells that can be observed and identified as the tumor. However, those cells can be aggregated into a bigger agent at a higher scale and lose the detailed level to see this tumor evolution over organs. The micro-level was useful to see and be precise on the creation of the tumor, once it has been created, this level is not useful in the model anymore and can, therefore, be removed.

Therefore, the choice of a pattern in the continuum to build an ML-ABM will have to be chosen by modelers following some modeling constraints or some more general software development limitations. In the first case, as in the previous example of the tumor model, modelers can choose a simpler construction of levels as in-between level emergence is not the main interest. Concerning software limitations, these can, for example, come from the fact that a model is composed of sub-models of different natures (ABM, EBM, etc.), which can be complicated to integrate into a Russian doll type architecture; in this case, it is simpler to use a Zoom or Collaboration type architecture. Also, the choice of an ML-ABM pattern can be a combination of both modeling and software choices.

Finally, integrating the human dimension as a dedicated level in an ML-ABM, as pointed out in Section 3, is the next big challenge of SES using agent-based models. Current solutions make it mostly impossible to escape the "unique hierarchy of viewpoints" to describe and simulate models. This prevents the implementation and coordination of multiple simultaneous viewpoints, which usually require multiple hierarchies (e.g., authors in the reviewing process) to define the functions and structures of a complex socio-environmental system.

In order to highlight the benefice of each approach as well as illustrate which questions can be answered by which multi-level coupling solutions, we propose to use a simple agent-based model of pedestrian evacuation (composed of `People` agents walking from left to right) along this review over every pattern presented.

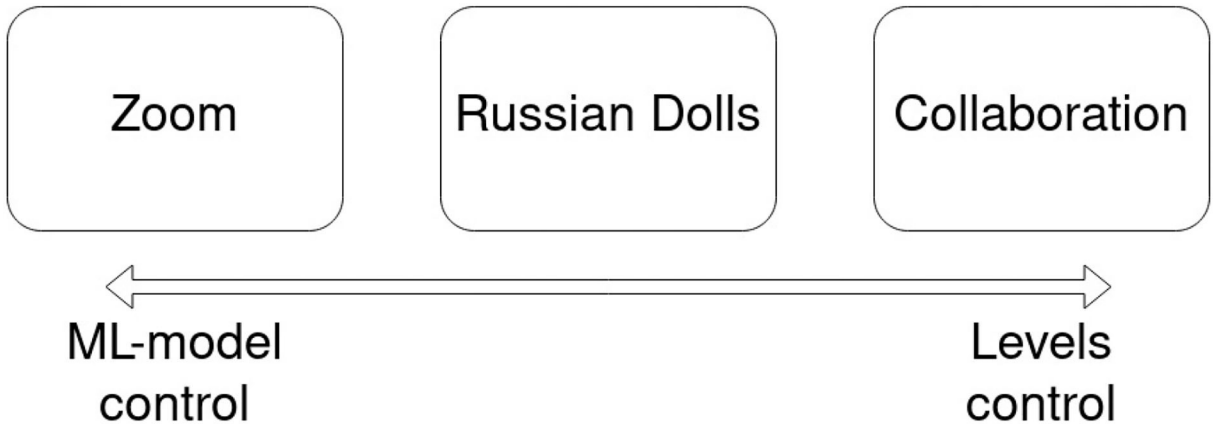


FIGURE 1
Multi-level agent-based modeling (ML-ABM) control's continuum going from control fully delegated to the global model in the Zoom approach, or keeping level's self-autonomy in the Collaboration approach.

2.2. Zoom

2.2.1. Problematic faced in SES

On the first end of the continuum, the most rigid control is the one represented by the so-called "Zoom" pattern. Each level is an independent model using a level of abstraction on the system using different kinds of models (agent-based, equation-based, else) with any scale. In this pattern, only a single level (i.e., model) is processed at a time, and modelers describe how the model will switch from one level to another using some transition function allowing to aggregate or disaggregate elements from one level to another.

In other words, while *zooming* in, the model will (1) call a transition function (defined by modelers) over the current level, (2) give this result to initialize the new level, then (3) destroy the previous level now unused. The system coherence over levels is insured by modelers with the transition function chosen.

However, as levels are not persistent in the execution of the model, the second SEs' problem of capturing systemic changes is very limited (or impossible), and the Zoom pattern is not used to answer this need.

2.2.2. Definition

To describe it in a more formal way: the Zoom pattern corresponds to a representation where levels are explicitly described (usually as agents) in the model with destructive behavior. Level behaviors, scales, transfer functions, etc. are also described explicitly. However, only one level is active at a time in the simulations. This ensures that there is no competition or conflict between the levels of representation during the execution of the model. Each level has an explicit aggregate and disaggregates transfer function that explains how to go from one level to the next one. Therefore, the use of this function will destroy the source level currently in use, to create the next one.

Figure 2 illustrates the global architecture of this pattern. In this illustration, the model has three levels: one equation-based model and two agent-based models. As shown, every level represents a different scale and is spatially (and temporally) limited by the level above. Finally, it is possible to move from one level to another, following some transfer functions (aggregation, disaggregation, etc.). This construction leaves each level independent, but only one level can be executed at a time.

If this Zoom pattern is used on the evacuation example model, it allows us to model and simulate the crowd either as an agent-based model (i.e., the initial model) or as a fluid mechanics equation (as in Henderson [23]). Zoom does not allow processing them at the same time; therefore modelers will have to define transition functions to move from one level to another. In this example, the mean-field approximation of agents [24] can be used to move from the ABM level to the EBM one as displayed in Figure 3.

2.2.3. Literature review

Differently from other patterns presented later, the Zoom implementation is simple and does not require a real framework to create this kind of ML-ABM. Most of the implementations are *ad hoc* (i.e., dedicated to the model it has been used for, and not generic to any model), then instead of listing them, this review will list how and which mathematical functions are used to allow switching from one level to another.

A note can nevertheless be made to highlight that this pattern is preferred in the modeling field of road traffic [25], as well as in crowd simulation [26, 27].

Some of the most commonly used transfer functions are aggregation approaches including mean-field theory [24, 28]

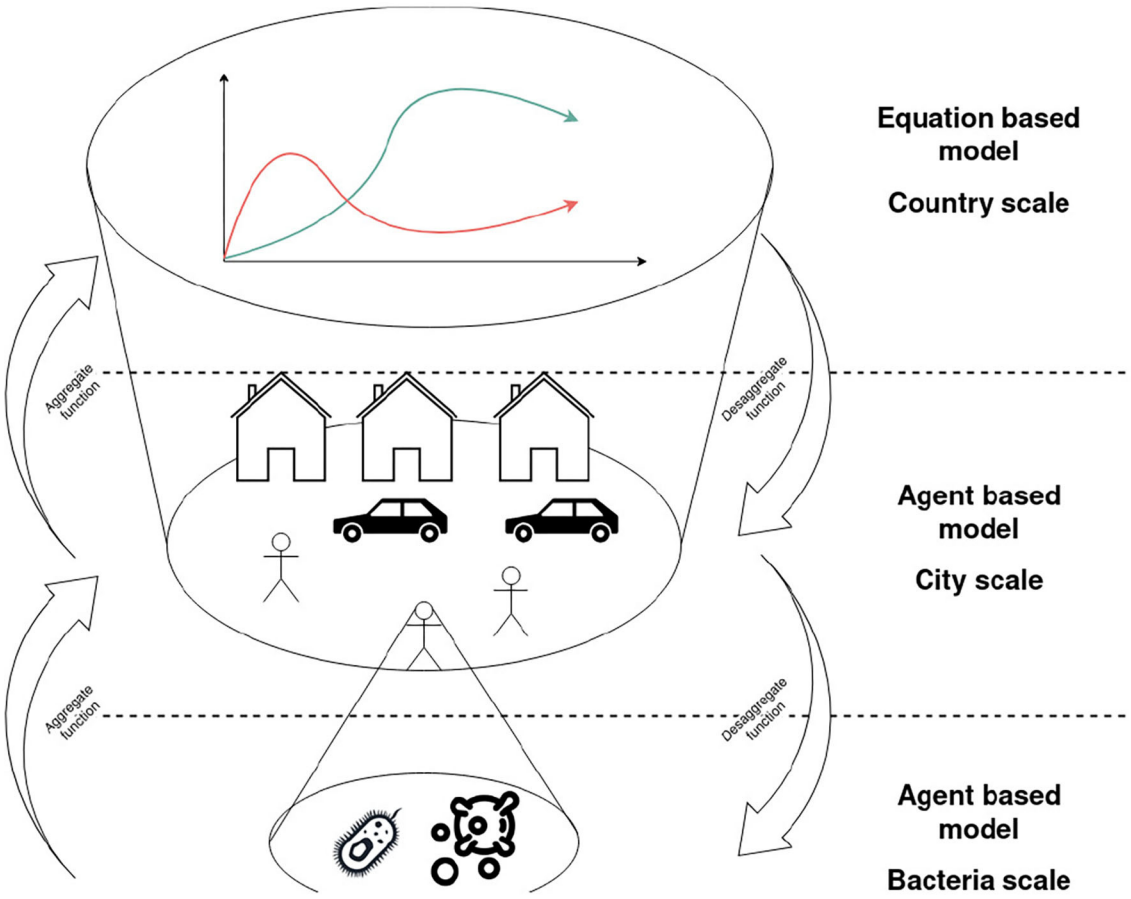


FIGURE 2
Simplified representation of the zoom architecture.

and aggregation of variable methods [29]. In those approaches, the studies of the behavior of complex dynamics at the micro level with many agent components interacting with each other and with the environment is approximated by a single average dynamics of the system at the macro level. Those approaches have been applied to a wide range of fields such as physics, artificial intelligence, epidemics, ecology, biology, and game theory. However, they have the drawback of not keeping a detailed representation of the system, and, in particular, do not offer any flexibility in information transported between levels of the system due to the principle of mathematical functions.

One way to overcome the limitation is to use a graph-representation as an intermediate level which allows more flexible information transfer from different levels of a complex system either as bottom-up or top-down information flow through levels. This idea was introduced in Nguyen [30] and in some related and extended work [31–33]. This methodology was first tested with the theoretical case study in population systems/ecology; and later rapidly applied to fishery systems, epidemiology systems [34], soil systems [35], and waste management systems [36].

2.2.4. Strengths and limitations

It is possible to extract some common strengths and limitations from the use of the Zoom pattern by modelers.

The greatest strength of this pattern lies in its ease of use and implementation by modelers. This simplicity is found in the independence of the levels between them. Actually, they are only linked by transition functions to switch from one level to another. In fact, this independence leaves, among other things, the modelers free in the nature of the models that can be used as levels (agent-based, equation-based, etc.).

Moreover, due to this autonomy of the levels, the development of the layers of a multi-level model can be done in parallel by letting a multi-disciplinary team work on different models (in their nature, temporal and spatial scales, etc.) which can be easily joined in a Zoom-type ML-ABM. Also, the development and extension of a single level can be done

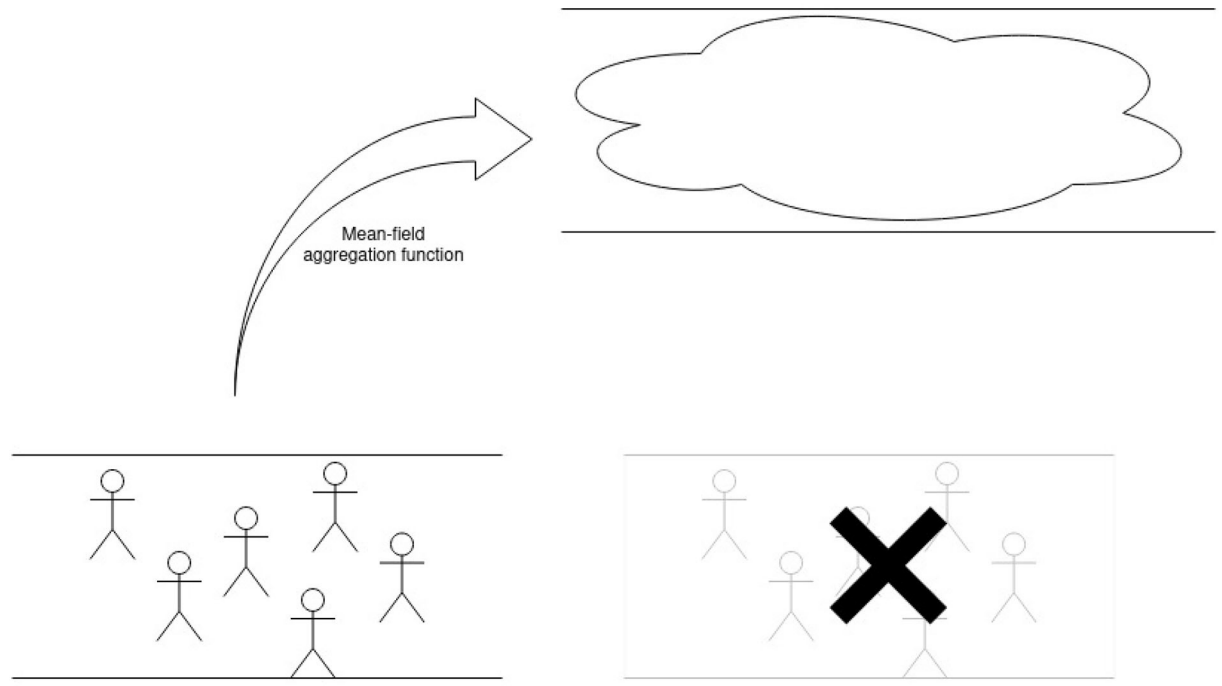


FIGURE 3
Simplified representation of the example model in the Zoom pattern aggregating agents from the micro-level to an equation at the macro-level. This operation is destructive, then after the aggregation, the ABM is removed and the model runs only with the macro-level.

independently of the global nesting of this level in the rest of the model.

Two more strengths result from the independence of the levels, namely that they are, by nature, reusable in other models and can be arranged at the discretion of the modelers. The reusability allows retrieving highly specialized models (ABM, EBM, etc.) from another team of modelers to improve and enrich the model currently developed. These models can be independent or come from another ML-ABM of the Zoom type. The choice in the arrangement of levels between them allows one to follow the vision and the architecture that the modelers wish to apply. It also reinforces the independence of the levels because they do not necessarily have to fit together according to scales or any other constraint (except for the feasibility of the transition function, which is virtually always feasible).

A final strength of this pattern lies in the resource saving it allows. Indeed, by only computing one level at a time, the addition of new levels will not impact the resources consumed by the execution of this multi-level model.

These forces create, nonetheless, a strong limitation in the use of this pattern. The destruction of unused levels (positive for the model resources' consumption) inevitably creates a loss of information from these levels and prevents the appearance and exploration of emergent phenomena at multiple scales and between these levels.

2.3. Russian Dolls

2.3.1. Problematic faced in SES

As a reminder, SES is using ML-ABM to try to deal with abstract representation at different scales (spatial and temporal) and to capture systemic changes in the whole model and in-between levels.

The Russian Dolls pattern tackles both problems by offering a less rigid control over the levels : they are provided with the possibility, when needed, to act on their own, but this follows a very hierarchical coordination. Each level is developed in a given scale and space, then they are all coordinated (i.e., scheduled as defined in ABM) together at the model point of view in regard to each scale. In other words, the temporal scale means that a level executed at a minute scale will be executed 60 times between steps of a level working at an hour scale; as for the spatial scale, each level is limited and bounded in a sub-space of the above level, like a house (sub-level) in a city (above-level).

As a result, this pattern forces levels to be used in a strict hierarchical order with each level smaller, or equal, in its temporal and spatial scale to the level above. Also, it is important to highlight that levels are no longer independent, making this ML-ABM very comparable to integrated models. Thus, compared to the previous Zoom pattern, no level (then nor information) is lost or destroyed which allows capturing the

emergence and systemic changes of each level of the model and in-between executed levels.

This architecture is based on a software engineering concept called "tight coupling" (or "strong coupling") [37]. A component (in ML-ABM, it is level) in this concept is usually highly dependent on each other and needs to know a lot about other components. Changing the internal logic of one component in a tightly coupled application often requires changes to several other components. But it offers the advantage of a very stable whole system and a strong integration from one component with the others.

2.3.2. Definition

This pattern, as well as its corresponding implementations in different platforms, offers support for representing multiple levels simultaneously. However, it must respect the constraint of strict control over how the levels behave and exchange information. This can be implemented using scheduling schemes similar to those in the holonic formalism [38].

In this holonic-like approach, levels are globally rescheduled to execute the full model by level (i.e., one level is executed at a time and the whole model is considered as a simple single-level ABM). It requires mixing processes and structures (which are central to systems thinking). The organization on multiple levels impacts both the structural and functional aspects of the model. First, in the structural aspects, elements (i.e., agent, level, etc.) belong to others, or are contained in others; second, in the functional aspects, elements depend on others' dynamics. These generally correspond to an implementation of concepts of spatial and temporal scales (playing a central role in multi-level mechanisms), and the extension or modification of some properties from ABM: agents now can "belong" to others or be contained in others, which makes those agents being "executed" by others and lost their own control.

In the example illustrated in Figure 4, the ML-ABM rescheduled following the Russian Doll pattern will be executed as follows. The macro-level (i.e., the city scale level) is executed where cars are moving according to a traffic model. Then, each building will be executed at the meso-level, calculating its energy consumption. Finally, for each room in the building, the micro-level is scheduled and runs each agent following another model. Once every level has been executed, the model starts a new cycle, following this same process.

Furthermore, each level can have a reciprocal effect within and with other levels, leading to some emerging patterns in the multi-level model. For example, the meso-level could be influenced by actions done at the micro-level: if a room is cooking, doing a videoconference, or reading a book, this will influence differently the energy consumed at the meso-level. Moreover, if we were to add a pollution model at a higher level than the current macro-level, it could be fed by our macro-level (with car's pollution) and meso-level (with electricity's generation pollution).

The global scheduling explained in the example can be re-adapted by modelers depending on their needs. It can start from the smallest level to the biggest (some Bottom-Up scheduling), or the opposite (Top-Down scheduling). Thus, all levels are synchronized and computed together, keep their agent's details, and can dynamically draw emerging properties in between levels.

This holonic implementation is very similar and comparable to what is called "tight coupling" (or strong coupling) in software engineering. It is a type of coupling which describes a system in which software parts are not only linked together but are also highly dependent on each other.

In this comparison, Russian Dolls' parts are its levels and share many developmental characteristics with parts of tightly coupled systems, which are often seen as disadvantages:

- A change in one module (level) usually forces a cascade of changes in the other.
- The software architecture composition of the modules (levels) might require more effort and time due to the increased inter-module dependency.
- A particular module (level) might be more difficult to reuse and test because dependent parts must be included.

Coming back to the case of the pedestrian example model, the use of the Russian doll architecture allows for example to integrate of this model, which allows simulating pedestrian movements on a single road, in a larger scale model simulating for example the traffic in the whole city. This coupling would benefit the bigger level to have a more precise simulation of critical points with a finer simulation of people evacuating an area. Those two levels would, however, lose their self-execution scheduler and be considered as a whole by the multi-level model. This way, considering that the city (i.e., big-scale level) has a step of 1 h, and the evacuating crowd (i.e., small-scale level) of 30 min, the model would execute the crowd one extra time between each city's execution as displayed in Figure 5.

2.3.3. Literature review

Different tools and frameworks have been developed over years to ease this kind of ML-ABM by modelers. Some of them have been made and used in the field of urban traffic as **Multi-Level Mesa** [39] provides methods to help manage the complex interactions of agents and modules of agents (e.g., groups) across multiple hierarchies (i.e., levels).

But, some more generic platforms have also been created for modelers either to apply some formalism or to extend already existing ABM platforms like:

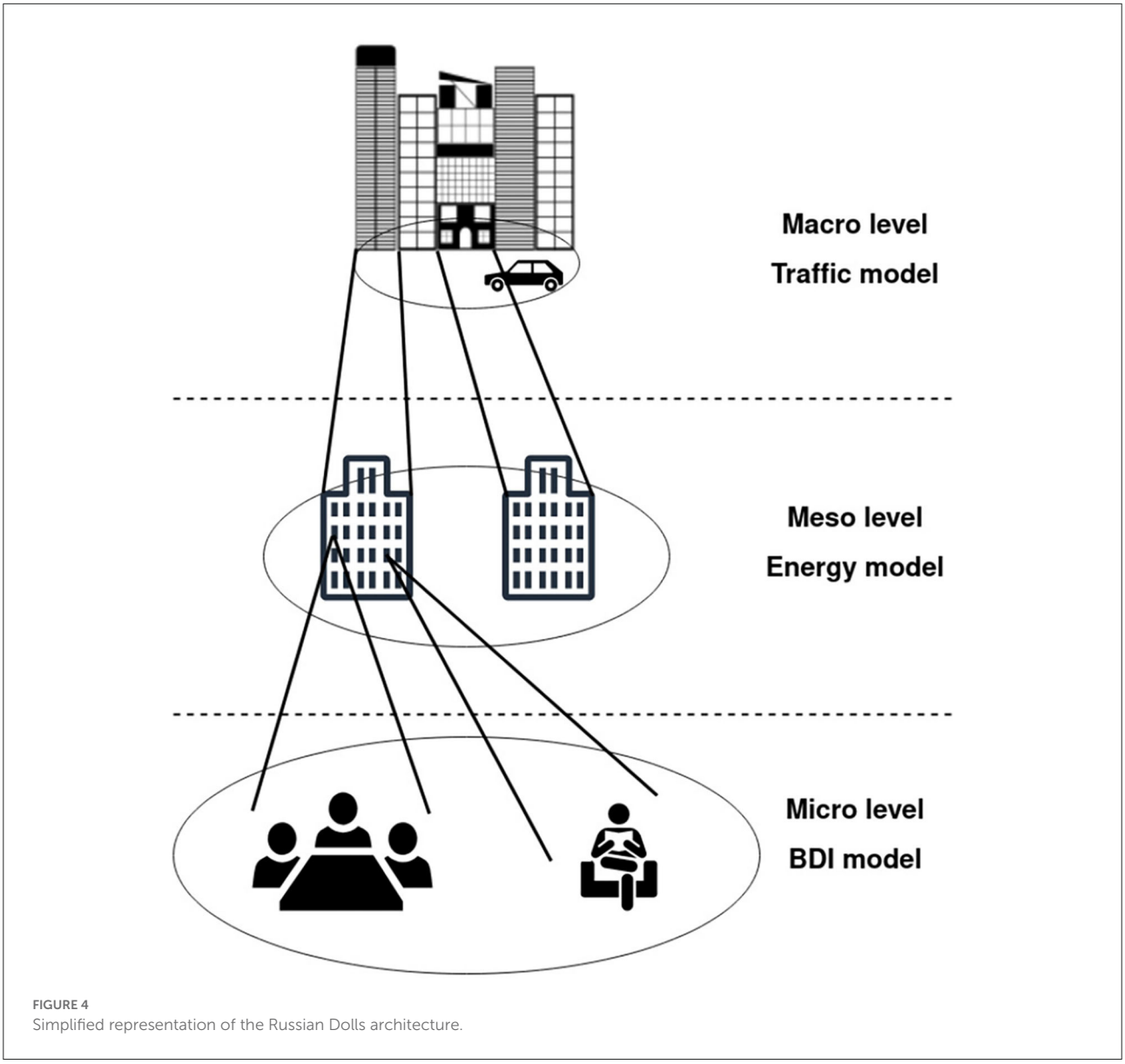


FIGURE 4
Simplified representation of the Russian Dolls architecture.

- **GEAMAS** (GEneric Architecture for MultiAgent Simulation) [40] is a pioneering ML-ABM framework integrating three levels of description (micro, meso, macro). The micro and macro levels represent, respectively, agent and system points of view, while the meso level represents an aggregation of agents in a specific context. Levels communicate with each other asynchronously. It has been applied to a variety of fields as thermodynamic models [41].
- **ML-DEVS** (Multi Level DEVS) [42] is an **extension of DEVS formalism** [43] which allows the simulation of multi-scale models (and not only coupled models in which the behavior of a model is determined by the behaviors of its sub-models). There are two types of relation between levels: information propagation and event activation. However, ML-DEVS focuses on multi-scale modeling and therefore, only supports pure hierarchies of models. This extended formalism has been implemented in the framework of James II [44] to develop multi-level models.
- **CRIO** (Capacity Role Interaction Organization) [45] is an organizational meta-model dedicated to ML-ABM based on the **concept of holon**. It has been specially developed for multi-scale simulations of pedestrian flow. It had been used to model human activities [46] as well as a base to build normative holonic metamodel for multi-agent systems [47].
- **LevelSpace** Hjorth et al. [48] is an **extension for NetLogo** [49] which allows modelers to dynamically create models

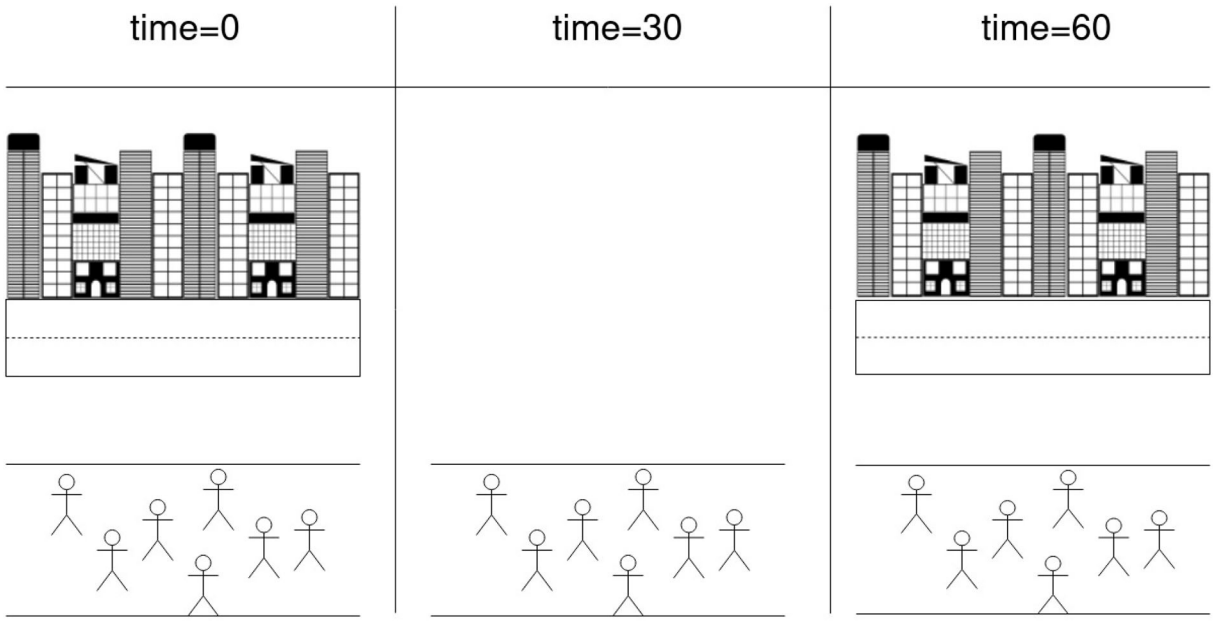


FIGURE 5
Simplified representation of the example model in the Russian Dolls pattern rescheduling both models' execution. The micro-model is processed alone in *time=30* as it has a smaller time step than the macro-model.

inside other models (in the same way as agents are created inside models), which implies a strong hierarchical structure.

- **Capture Release** Vo [18] is a **GAMA Platform extension** [50] that allows aggregate agents in the above level's agent. This process changes the behavior of captured agents which will be controlled by the bigger agent. It has been used, for instance, on an evacuation model [20] and several toy models displaying the clusterization of boids [51] as flocks [52].

Finally, some *ad hoc* implementations have also been used in different projects and research fields with similar approaches as in medicine with some tumor modeling [53], in urban traffic [54, 55] or even in hydrological modeling project [56].

2.3.4. Strengths and limitations

It is possible to extract some common strengths and limitations from the use of the Russian Dolls pattern by modelers.

Level's persistence is one of the strengths of this pattern in regard to the fifth SES challenge from Elsawah et al. [8]. It allows modelers to analyze and capture emergence between levels and explore modeling problems over multi-level representations of the system.

Another strength of this architecture is its natural hierarchical architecture. The spatio-temporal imbrication of levels is what is perceived from our reality and makes the conceptualization of those multi-level models simple.

However, this strength also creates one of the major drawbacks of this pattern as levels are not independent, which creates several problems shared with tight coupling in software engineering. As levels are integrated into the full model, extending or debugging them becomes more complex.

This design pattern is pretty close to classical ABM as it does not bring any new concept or method to build or use ML-ABM but extends some (like ABM's scheduling) for a multi-level explicit purpose. Therefore, it has been very straightforward for the modeling community to develop this pattern and create a wide variety of simple and powerful tools to build Russian Dolls.

But, this strength also creates a limitation in the nature of usable levels. As the full model will be close to a simple ABM, it is difficult to add levels that are not agent-based (but not impossible as in Chapuis et al. [57]). Furthermore, it will be impossible to use anything else than an ABM for an intermediate level (like in a multi-level model architecture with some ABM at the macro level, equation-based at the meso, and ABM at the micro). This kind of structure is possible with the last pattern (Section 2.4).

2.4. Collaboration

2.4.1. Problematic faced in SES

As a reminder, SES is using ML-ABM to try to deal with abstract representation at different scales (spatial and temporal) and to capture systemic changes in the whole model and in-between levels.

At the other end of the continuum, we find a pattern of Collaboration that represents a loose control between autonomous levels. Each level is, therefore, considered as an independent model using any level of abstraction on the system using different kinds of the model (agent-based, equation-based, else) with any scale.

However, to ensure catching emerging patterns in and in-between levels, those are not destroyed and created during the model's simulation run. Similar to the Russian Dolls, they all are collaborating and processed at the same time.

This architecture derives from software engineering concepts called "loose coupling" or "weak coupling" (and is opposed to "tight coupling") [58]. The idea behind this concept lies in two main aspects: 1) components (for ML-ABM it is level) are weakly associated (i.e., have a breakable relationship) with each other, making changes in one component least affect the existence or performance of another component; and 2) each component has little (or no) knowledge of the definitions of other separate components. Those points protect level autonomy and independence, especially the second one preventing levels to lose their own scheduling/execution control.

The drawback of this flexibility is the complexity of the use of such a pattern. Even though this kind of coupling has been experienced for years and is widely adopted by the software development field, it requires some technical software work to connect each level together with some unusual concepts for modelers.

2.4.2. Definition

Similar to the Russian Dolls pattern, the last pattern, Collaboration, also addresses some modeling needs which are not addressed by Zoom. It keeps the level's information from being lost. However, to ensure the independence and self-control of each coupled level, this pattern applies a loose coupling (or weak coupling). Therefore, the main difference between the second pattern and Collaboration is how they both solve the coupling problem between levels.

Collaboration allows us to get rid of the global rescheduling constraint from the previous pattern. Here, the architecture is coupling independent models as levels and allows reciprocal information exchange between them. As this interaction is bidirectional, levels have a direct influence on each other.

In computational and systems design, a system is said to be loose when one component has a breakable relationship with other components, and components have no knowledge (or little) of the definitions of other separate components in the system. In the Collaboration pattern, components are model's level which are all communicating by software messages and considering each other level as a black box. Moreover, since the levels are loosely coupled and considered a black box, each level can be seamlessly replaced by another model. It is also possible to extend a given Collaboration model with another pre-existing model as in Chapuis et al. [59].

In Figure 6, each level corresponds to an independent model that sends information (the output result of each processed cycle) to other levels (as the input value for the processed cycle) in directions chosen by modelers. Those messages are not limited by the hierarchical order of levels in the model, e.g., the micro-level can directly interact with the macro-level, and reverse.

Therefore, by changing the way levels are coupled, this pattern no longer requires modelers to focus on or work on level scheduling. Because each level is considered a black box, they are all scheduled by themselves, each keeping control of the evolution of the model's level. This control distribution among the levels is in particular interesting in terms of the representation of systems where functions are, therefore, also distributed.

However, this distribution comes at the cost of a greater complexity in the models, as the interactions between levels are less structured, and all the related problems of coherence between levels can raise.

In the example evacuation model (Figure 7), it is possible to apply this pattern in order to evacuate the crowd on two exit points respecting a fluid physics model [60]. Therefore, the modeler will create an interface allowing to send output data (e.g., `Person` agent entering this area) which will be taken as input by the second model. That second model will process its fluid simulation then send back information on which path agents are taking (Figure 7). This way, instead of representing the crowd as agents making a choice, the modeler can consider it as a fluid that will force agents to one exit following a more macro-scale crowd dynamic.

2.4.3. Literature review

Different tools and frameworks have been developed over years to ease this kind of ML-ABM by modelers. Some of them have been specifically made and used in the field of urban traffic as

- **SimMobility** Lu et al. [61] is structured in three components (long-, mid-, and short-term) and follows a multi-level approach based on the time aspect. Modeling aspects are distributed across the three components and brought together into a single database.

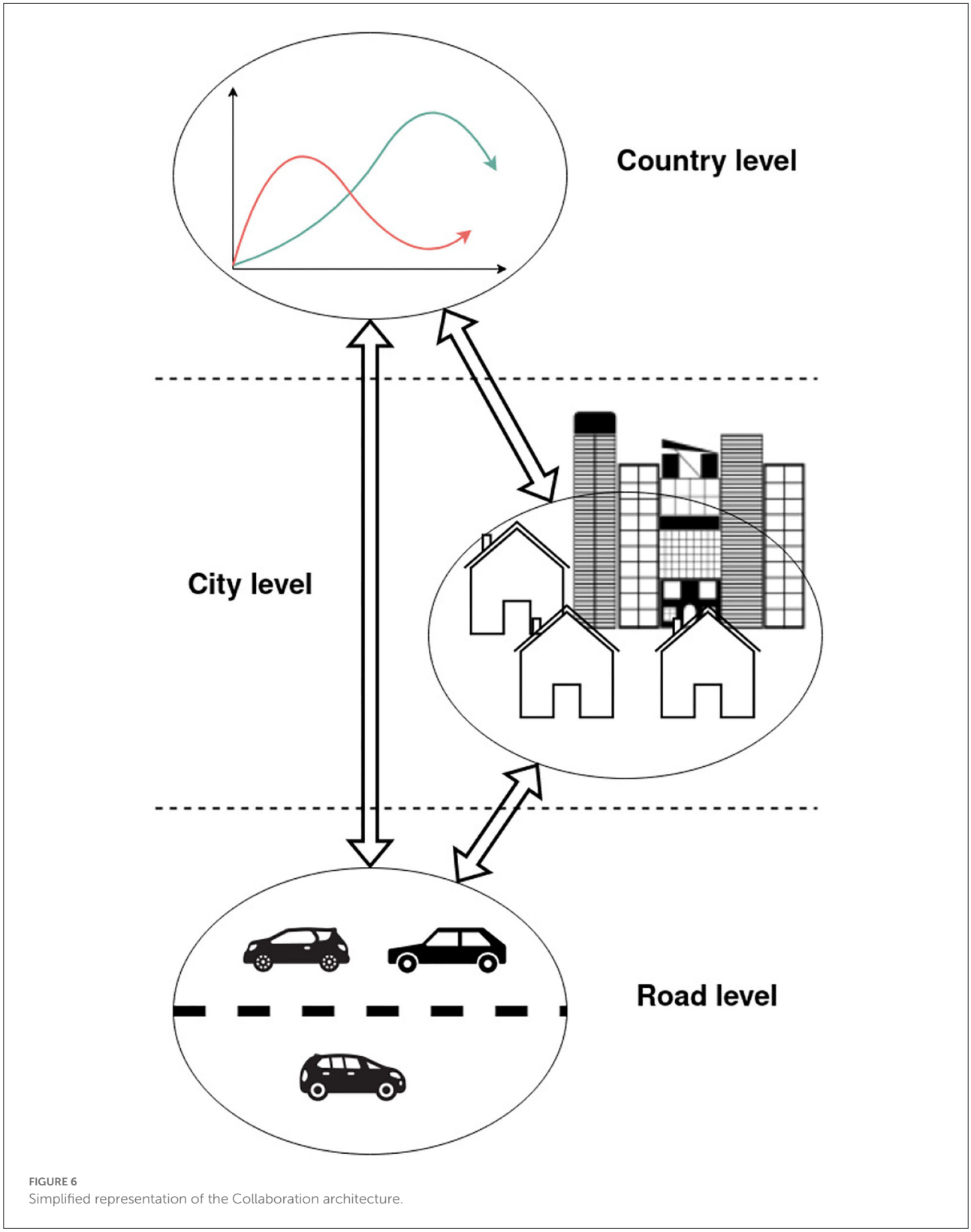


FIGURE 6
Simplified representation of the Collaboration architecture.

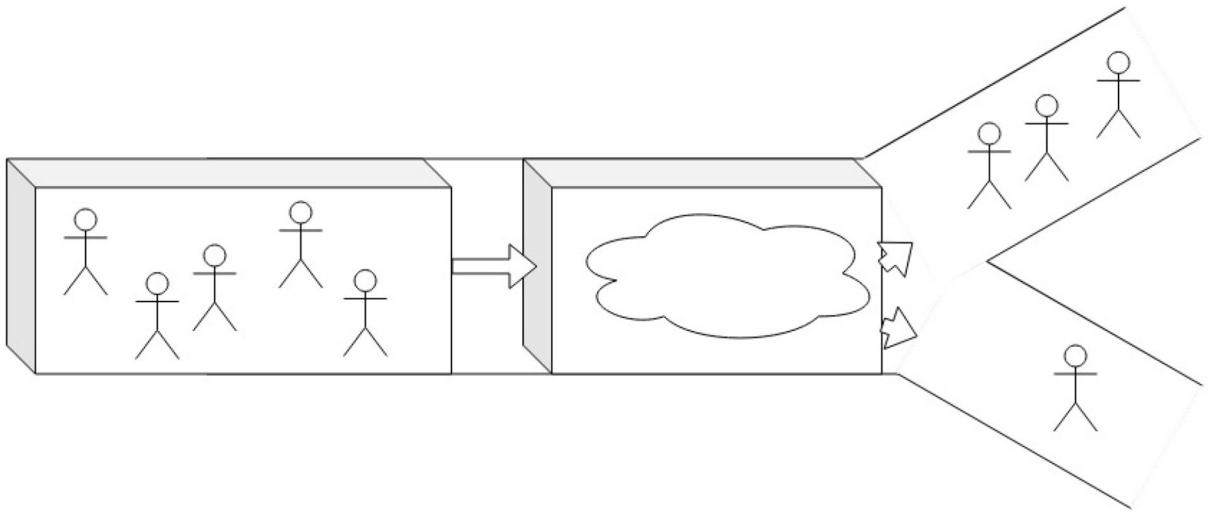


FIGURE 7
Example of the *Collaboration* pattern allowing to simulate the example agent-based evacuation model than relying on a fluid model simulating the crowd as a fluid to process the number of agents going to one exit or another. Those models are processed independently and are considered by each other as black boxes.

- **TraSMAPI** (Traffic Simulation Manager Application Programming Interface) [62] is a system architecture integrating (with weak coupling) SUMO [63] and JADE [64] frameworks to build ML-ABM. JADE agents represent drivers (managing micro-behavior such as lane changing) that are linked to vehicles in SUMO (managing meso travel modes).

Some generic tools have also been created for modelers either to apply some formalism or to extend already existing ABM platforms like:

- **GEAMAS-NG** David et al. [65] is an updated version of Russian Dolls' GEAMAS, where the framework provides tools to detect and reify emergent phenomena between different kinds of models (like MAS and CBR [66]). It has been used for modeling urban area evolution.
- **Co-modeling** Huynh [67] is a **GAMA Platform extension** changing the base concepts of ABM formalism by allowing agents to be models themselves. With it, the ML-ABM consists of a collection of models that are only loosely coupled. It has been used to couple an ABM with a hydrological model in an evacuation model [59].

Finally, as for other patterns, some ML-ABM has been developed using this design pattern but without any framework in an *ad hoc* implementation. It is the case in fields such as biological systems, e.g., in many works on the multi-level modeling of tumor development [68, 69], urban traffic [70] or resource management models [71]. An interesting usage is made by Yang et al. [72] who use the flexibility of this pattern to easily swap the micro level and explore its impact without having to modify the rest of the model.

2.4.4. Strengths and limitations

These pattern offer qualities from both previous patterns. Similarly to the Zoom pattern, it allows to use and reuse of any kind of independent model as a level in the built multi-level model, and, as with the Russian Dolls, it preserves levels during the whole model's execution allowing the exploration of emergence in a multi-level system. Additionally, due to its loose coupling base, it is possible for modelers to conceptualize and coordinate levels in a hierarchical order of their wish.

However, this pattern is also the most complicated for non-developer scientists and modelers to use. For example, the comodeling framework requires three extra technical files to ensure communication and coherence between two levels (making a total of five files). This creates a great deal of complexity when trying to develop models.

Also, its overall complexity lets modelers precisely explicit the coherence of levels' execution, leading to greater complexity in the development and can make multi-level wrong because of it.

Because of this overwhelming cumulative complexity, few tools currently exist that are capable of hiding this complexity

and allowing a wide range of modelers to use and explore this pattern.

3. Discussion

3.1. Limitations in the reviewed continuum

Each of the approaches and ML-ABM architecture along the continuum presented in this review have demonstrated its utility, usability, and interest. They all have been used by modelers in different fields to answer different scientific questions and allow, different, to model complex systems in the ABM model over several levels of abstraction. Therefore, they all achieve the same goal to allow modelers to create, experiment on, and work with multi-level models.

However, they also share a strong limitation in regard to SES requirements: None of them propose ways to explicitly represent spatial or temporal scale within or between levels. This limitation is induced by the use of ABM in which those concepts are missing.

Thus, the Zoom pattern simply forgets this issue between levels by considering only one structural and functional level at a time. The spatial scale can, in a way, be expressed in the transfer function between levels, but the temporal scale is always induced in the model and modelers are responsible for respecting this coherence between levels of the model they are building. The two other patterns, Russian Dolls and Collaboration, do have some workarounds to address those needs, though they are limited. The temporal constraint and coherence can be translated with scheduling and the spatial constraint is translated by the encapsulation of levels in each other.

Those solutions allow us to address and translate these spatial and temporal constraints between levels in a somewhat limited way, although they are sufficient in most cases. For example, complex system models following the natural single hierarchical imbrication, like in urban traffic models [54] or biological models [53], totally fits with implicit this spatio-temporal scales in ML-ABM; but some SES models do not (e.g., the scientific reviewing process).

Since SES are interested in the agent's behavior, which can be used as a level but not indexed in any natural spatio-temporal hierarchy, it requires the possibility to explicitly define those constraints. This point reveals another strong limitation shared by each of those patterns: they are all restricted to represent only one hierarchy at a time. It forces modelers to mix functional and structural properties in describing levels and forces them to express only a single hierarchy, despite the fact that several hierarchies of points of view and abstraction may be necessary to understand the functioning of a complex system and, moreover, socio-environmental ones.

This limitation prevents modelers from working on a wide range of models. We can illustrate this with EasyChair's organization: how is it possible for the same person (i.e., agent) to be an author, reviewer, and member of the program committee, at the same time? Taking back this review's example, evacuating agents can, in the real world, also have several roles as *Pedestrians* with the only will to go to the exit point, *Parents* varying this first behavior with a will to keep their children close and safe, *Police Officer* trying to evacuate everyone before himself, or a combination of each. How can it be simply implemented by modelers over the pre-existing model, and how can modelers choose which role have the greatest influence on the agent? Over which conditions?

Each of those roles brings a single physical agent to have different behavior (when it is an author, it will listen to comments and try to please reviewers, but this comportment will be entirely reversed when the agent is wearing the reviewer role), interaction range (interacting with other agents from the same group with the same role), etc. This pictures the limitation of models allowed by the use of a single hierarchy and, if used, needs to over simplify the model in a wrong abstraction.

3.2. Toward a more faithful representation of multiple-hierarchy in ABM

Although ABM does not really provide a way to represent the entirety of the concepts used in complex system science (notably simultaneous hierarchies of viewpoints), its software basis, multi-agent systems (MAS), has explored some interesting paradigms that could be used to enrich the patterns presented above.

Multi-agent systems alone provides a good basis for building complex agent-based systems, but it is one of its extensions, the organization centered multi-agent system (OCMAS), which can serve as a real base to address current SES ML-ABM drawbacks.

The concept of "organization" is hard to define, but Ferber et al. [73] described its main features as follows:

- An organization is constituted of agents (individuals) that manifest a behavior.
- The overall organization may be partitioned into groups (partition) that may overlap.
- Behaviors of agents are functionally related to the overall organization activity (concept of role).
- Agents are engaged in the dynamic relationship which may be "typed" using a taxonomy of roles, tasks, or protocols, thus describing a kind of supra-individuality.
- Types of behaviors are related through relationships between roles, tasks, and protocols.

An important element of organizations is the concept of role. A role is a description of an abstract behavior of agents. A role describes the constraints (obligations, requirements, skills) that an agent will have to satisfy to obtain a role, the benefits (abilities, authorization, profits) that an agent will receive in playing that role, and the responsibilities associated with that role.

A role is, then, a description of patterns of interactions that an agent will have to perform. Organization can be considered at two different levels: at the organizational (or social) level (i.e., as a composition of roles) and the concrete (or agent) level (e.g., a crowd) [74].

Another important principle described by Ferber et al. [73] is that the organizational level describes the "what" and not the "how." The organizational level imposes a structure onto the pattern of agents' activities, but it does not describe how agents behave. In other words, the organizational level does not have a "code" that can be executed by agents, but it provides specifications using rules and expectations placed on the agents' behavior.

3.2.1. Agent, group, and roles

One interesting approach to treating this is the **Agent/Group/Roles** (AGR) paradigm [75]. When used, it is possible to extend the behavior of the agents belonging to roles that are changing or adding behavior to their existing ones. The AGR formalism is, thus, defined as:

- **An agent** is defined as an active communicating entity that takes on roles in groups. An agent becomes a member of a group by taking on a role defined by this group. It can then simultaneously take charge of other roles of this group, or roles defined in other groups, of which it then also becomes a member. No assumption is made about the architecture of the agents.
- **A group** consists of a set of agents interacting through their roles. At an abstract level, it defines a set of roles and relations between these roles that realize a global function. A group is then instantiated by the assumption of its roles by an agent, but all the roles in it may not be assumed, and the same role may be instantiated several times on different agents. The group is, therefore, a description structure with a very high level of abstraction and is capable of generating a great diversity of realizations. Finally, groups are opaque to each other: agents can only interact if they belong to the same group.
- **A role** is defined as the abstract representation of the function of an agent in a group. Roles are defined in groups and encapsulate how an agent should act within the group.

To allow the application of the AGR principle, an in-depth re-work of how Agent-Based Models conceptually (following the AGR paradigm) and technically (in how frameworks and software tools implement models) work is needed. Indeed, the current structure of the ABM is based on the composition of agents that are owned and compose groups. The architecture of ML-ABM is based on this composition of the group of agents.

Following the new AGR structure, it needs to be changed to allow agents to be owned and fed by different roles. Then, those roles could compose groups on which it would be possible to re-implement the current ML-ABM design (Figure 8).

This modification would deeply change the way levels are composed. They will no longer be spatially bounded to the agents composing them, and will instead let modelers create levels that rely on non-hierarchical, non-spatial, and non-temporal interactions between agents and roles. This new architecture allows the creation of complex SES models as the EasyChair's organization previously (Section 3.1) described in Figure 9.

In the AGR paradigm, illustrated in Figure 9, a person (i.e., agent) can have several roles, so it can be an author, a reviewer, and a member of the program committee. Each role would cumulatively enhance the behavior of the agent. Agents can also be grouped by role allowing committee members to communicate together and know other members of the group, which allows removing any spatial need constraint (from the agent and the group).

This new structure allows models to escape the single vertical hierarchy between levels. In Figure 9 1) levels are neither spatially nor temporarily limited by other ones, and 2) several levels can operate at the same hierarchical level as different evaluation groups in the diagram.

3.2.2. AGR review

Several frameworks have been proposed following this new architecture proposition, including some by J. Ferber himself:

- **AALAADIN** Ferber et al. [76] is the first meta-model of an artificial organization implementing AGR created by J. Ferber. It allows for building multi-agent systems with different forms of organizations such as market-like [77] and hierarchical organizations as well as in general modeling like in complex industrial process modeling. This first implementation has been later extended to a generic Java library called **MaDKit** [78, 79].
- **ORIGAMI** (Original Roles Identification, Groups and Agents Modeling Itinerary) [80] is based on the use of UML diagrams to build organizational representations of a system based on the AGR formalism. It allows structuring the analysis of a system through the identification of its

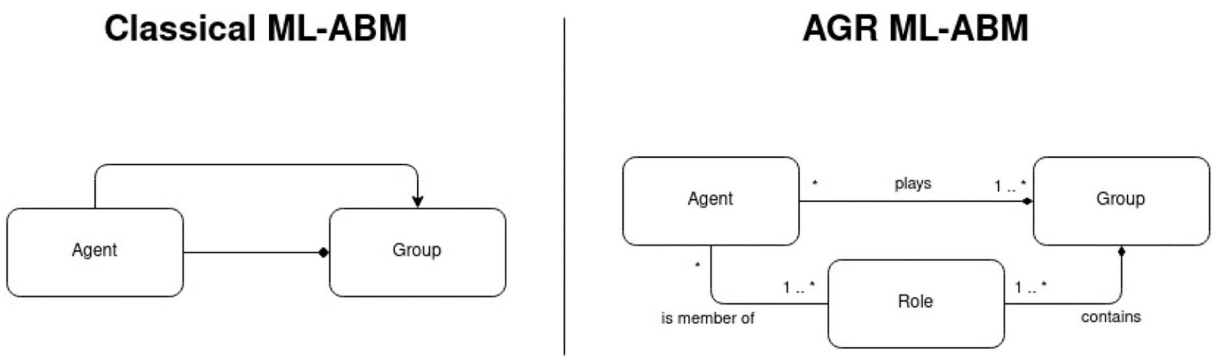


FIGURE 8
Classical ML-ABM UML **(left)** compared to a simplified AGR paradigm UML **(right)** [76].

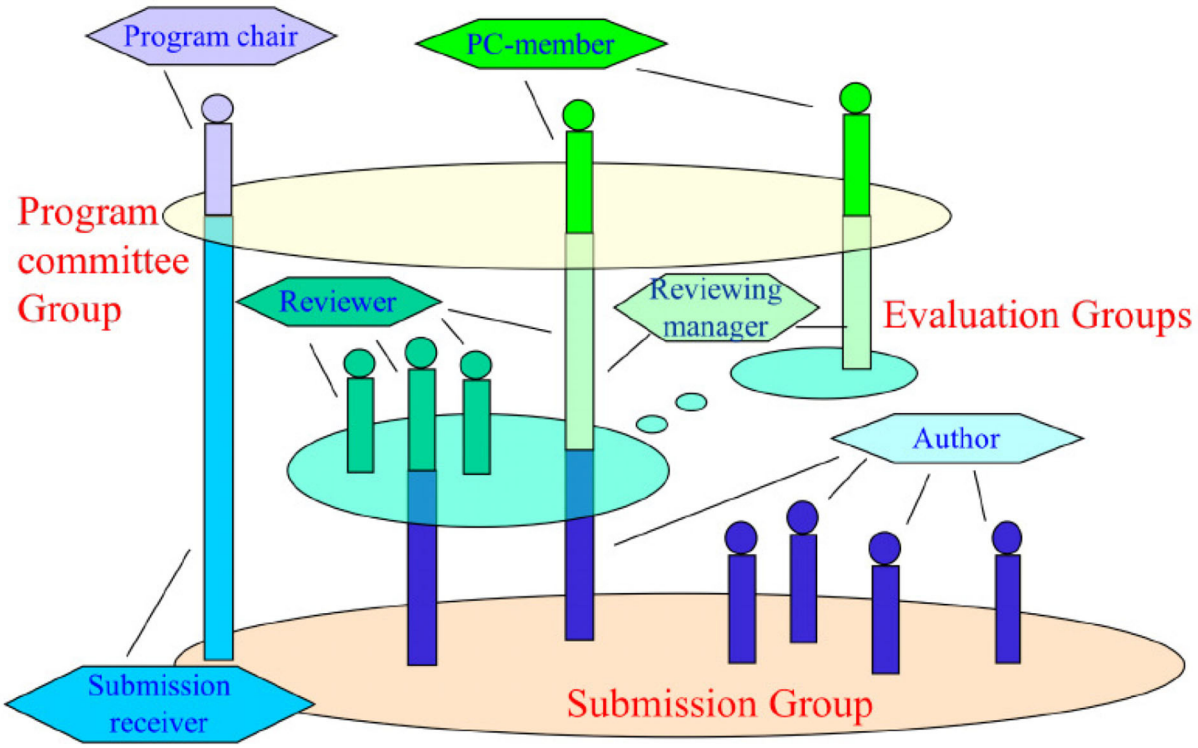


FIGURE 9
The cheese-board diagram describing the EasyChair's organizations in AGR [76].

structural units, the functions described, and the entities that constitute it with the use of diagrams. It has been demonstrated on a toy model reimplementing a collective management model of water irrigation system [81].

As the AGR formalism deeply change the structure of (ML-)ABM, some frameworks aims to extend it to make it more compliant with classical ABM and avoid having to change it:

- **AGRE** (AGR + Environment) [82] is an extension of the AGR organizational model, which includes physical (or simply geometrical) environments called "area." This extension is based on the concept of a space which can be seen either as a physical area or as a social group.
- **IRM4S** (Influence Reaction Model for Simulation) [83] is an ABM meta-model based on the Influence Reaction

model which views action as a two step process: first, agents produce "influences" (i.e., individual decisions) according to their internal state and perceptions, then, the system "reacts" (i.e., computes the consequences of influences) according to the state of the world [84]. The relations of perception and influence between levels are specified with digraphs.

In its presentation paper, this framework have been demonstrated on robot soccer collaboration model and the predator prey model [83] and reused as a base for a smart traffic simulator [85].

- **IRM4MLS** (Influence Reaction Model for Multi-Level Simulation) [21] is a multi-level extension of IRM4S. It relies on a generic vision of multi-level modeling. Therefore, interactions between levels are not constrained. It has been successfully applied to some real cases transportation systems [86].

Both frameworks aim to address the limitation discussed in Section 3.1 of ML-ABM, which does not fit some new multi-level models. They all extend agents with the new concept roles. Roles can be used in two directions: 1) to describe groups (which can be used as levels) and 2) to describe the aggregation of roles to describe agents. This model's structure enhancement removes the constraint to have agents belonging to only one super-level (as roles describe levels and one agent can belong to several roles) but also creates greater complexity.

Unfortunately, none of them is actually implementing AGR or proposing convincing tools for modelers to use and apply this innovative paradigm. Regardless of their usability, each of them tried to fit the AGR paradigm into either an existing framework or the general ABM architecture. However, as discussed in this review (Section 3.1), using both methods fundamentally cannot work. Then, most of the reviewed frameworks do keep spatial and temporal constraints between levels and failed to tackle the first described ML-ABM limitations.

Therefore, research and experimentation into tools that would allow modelers to *truly* use the AGR formalism is a topic of interest and could lead to some great new models.

4. Conclusion

This review explored how agent-based multilevel modeling (ML-ABM) has evolved and what solution it has offered to meet the representation needs of modelers. Three high-level architectures (called Design Patterns in this review) were described as landmarks on a continuum that allowed us to classify ML-ABM tools and frameworks for modelers from very simple approaches giving control to only one level at a time to more complex approaches allowing for the representation of relatively sophisticated feedback loops between autonomous levels.

The first landmark, called (Section 2.2), is the easiest to implement and allows modelers to move from one level to another with so-called "destructive" disaggregation/aggregation functions. It has the advantage of being simple to implement for the modelers and can be used with any type of model, but it only allows representation of one active level at a time, which limits the possibilities to represent and focus on the emergence relations between levels. It also presupposes a complete knowledge of how the different levels are implemented, neither allowing reuse of legacy models nor imagining separating them to make autonomous models.

The progression along the continuum is logically done in terms of the autonomy of the individual levels; the second landmark, which we call Section 2.3, marks the possibility of describing levels that can be active simultaneously but organizing them within a rigid holonic structure, where time and space scales are necessarily hierarchically nested. It allows modelers to describe the levels in a relatively natural way at the cost of a certain rigidity of the whole and a lack of autonomy of the individual levels as models.

The last one, which we have called Section 2.4, is the most complicated to implement for the modelers, but it offers the most flexibility in the description of the relations between levels. Based on the principle of "weak coupling" from software engineering, this approach considers each level as a black box that communicates through predefined inputs and outputs, which allows describing their interactions and feedback loops in a very free way. The strength of the approaches that follow this pattern is that they do not lock the modeler into a too rigid approach, and allow much better reuse of existing models. Their weakness, on the other hand, is that the design and validation of the models are made much more complex.

It is important to emphasize that this continuum is only used to classify the studies we have reviewed. It obviously does not represent a qualitative value that would express that some multilevel approaches are "better" than others, as this can only be judged in relation to the objectives of the modelers wishing to answer a specific question in their representation of SES. In some cases, simply capturing the transition from one level to another (appearance of a structure or behavior causing the transition to a higher level model) is sufficient, and this is what the vast majority of published study is limited to. In other cases, cyclic behaviors of aggregation/disaggregation, emergence/constraint, or more complex feedback loops must be implemented between levels to allow modelers to answer the questions posed to them, and this is what approaches ranging from Russian Dolls to Collaboration allow.

In terms of SES representation, the review has nevertheless highlighted (Section 3.1) two important limitations shared by all the approaches presented: 1) the lack of convergence on a common and documented representation of time and space scales, which remains at the sole discretion of the modelers and is often not explicit, making the reuse of

models representing these levels quite problematic; 2) the impossibility of representing more than one hierarchy at a time, which prevents the implementation and coordination of several simultaneous points of view, sometimes necessary to define the functions and structures of a complex socio-environmental system, in particular when multiple social groups—and their overlaps and interactions—must be represented.

We hypothesize that as the questions posed to modelers become more complex, these two limitations will become real obstacles to a more professional use of ABM. On the one hand, because, as multilevel modeling becomes more widespread, it will be necessary to be able to easily reuse the models of levels, which will have to make their time and space scales explicit; on the other hand, because the possibility of taking into account and interweaving different points of view is one of the prerequisites of interdisciplinary approaches, which become widespread in the handling of SES problems. In order to make progress on these two points, we believe it is necessary to revise the classical structure of agent-based modeling, which is still too closely linked to the limitations of object-oriented programming, from which it is inherited, in order to offer a more faithful representation of the multiple levels and to link them more explicitly to scales of time and space. The use of the AGR paradigm (Section 3.2) is in this respect potentially interesting. It allows one to define an agent as a dynamic set of roles and to define the groups of agents (the levels) in terms of the grouping of these roles. An agent can thus be described as belonging to several distinct levels simultaneously. The conditions for aggregating, disaggregating, or sharing control are defined by the group itself, which also allows, in theory, to add or subtract them from the model dynamically and also to build, eventually, easily reusable libraries of levels.

This proposal implies modifying the "classical" metamodel of agent-based models while keeping their easily accessible side for non-experts and continuing to ensure that their implementation is not too complex or constrained. This is not an easy challenge to meet, as it is also this simplicity that has made them successful in recent years. But as agent-based models become indispensable tools in decision support processes, moving away from the "toy model" aspect that has often been reproached to them, it seems important to us to make sure that their development becomes more professional. We hope that this review, these proposals, and the study that we are going to do on them will allow us to progress in this direction.

Author contributions

AB was the principal designers of this document and is responsible for much of the writing. AD and DN-N brought some complementary references and helped to structure the whole review. All authors have worked together, each in their own specialty, to design, and describe the research presented in the paper.

Funding

This work was supported by the ANRS (https://anrs.fr) through the funding of a COVID-19 SOUTH project (COV15) and the IRD Vietnam (https://vietnam.ird.fr).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Crutzen PJ. Geology of mankind. In: Paul J, editor. *Crutzen: A Pioneer on Atmospheric Chemistry and Climate Change in the Anthropocene*. Berlin: Springer (2016). p. 211-5.

2. Giupponi C, Ausseil AG, Balbi S, Cian F, Fekete A, Gain AK, et al. Integrated modelling of social-ecological systems for climate change adaptation. *Socio Environ Syst Model*. (2022) 3:18161. doi: 10.18174/sesmo.18161

3. Berkes F, Folke C. Linking social and ecological systems for resilience and sustainability. Linking social and ecological systems: management practices and social mechanisms for building resilience. *Environ Dev Econ*. (1998) 4:237–42. doi: 10.1017/S1355770X99220165

4. Anderson P. Perspective: complexity theory and organization science. *Organ Sci*. (1999) 10:216–32. doi: 10.1287/orsc.10.3.216

5. Gil-Quijano J, Louail T, Hutzler G. From biological to urban cells: lessons from three multilevel agent-based models. In: *International Conference on Principles and Practice of Multi-Agent Systems*. Berlin: Springer (2010). p. 620–35.

6. Simon HA. The architecture of complexity. In: *Facets of Systems Science*. Berlin: Springer (1991). p. 457–76.

7. Serugendo GDM, Irit MP, Karageorgos A. Self-organisation and emergence in MAS: an overview. *Informatica*. (2006) 30:45–54.

8. Elsawah S, Filatova T, Jakeman AJ, Kettner AJ, Zellner ML, Athanasiadis IN, et al. Eight grand challenges in socio-environmental systems modeling. *Socio Environ Syst Model*. (2020) 2:16226. doi: 10.18174/sesmo.2020a16226

9. Gain AK, Hossain S, Benson D, Di Baldassarre G, Giupponi C, Huq N. Social-ecological system approaches for water resources management. *Int J Sustain Dev World Ecol*. (2021) 28:109–24. doi: 10.1080/13504509.2020.1780647

10. Drogoul A, Huynh NQ, Truong QC. Coupling environmental, social and economic models to understand land-use change dynamics in the Mekong Delta. *Front Environ Sci*. (2016) 4:19. doi: 10.3389/fenvs.2016.00019

11. Lippe M, Bithell M, Gotts N, Natalini D, Barbrook-Johnson P, Giupponi C, et al. Using agent-based modelling to simulate social-ecological systems across scales. *Geoinformatica*. (2019) 23:269–98. doi: 10.1007/s10707-018-00337-8

12. Kelly RA, Jakeman AJ, Barreteau O, Borsuk ME, ElSawah S, Hamilton SH, et al. Selecting among five common modelling approaches for integrated environmental assessment and management. *Environ Model Software*. (2013) 47:159–81. doi: 10.1016/j.envsoft.2013.05.005

13. Schulze J, Müller B, Groeneveld J, Grimm V. Agent-based modelling of social-ecological systems: achievements, challenges, and a way forward. *J Artif Soc Soc Simulat*. (2017) 20:3423. doi: 10.18564/jasss.3423

14. Groeneveld J, Müller B, Buchmann CM, Dressler G, Guo C, Hase N, et al. Theoretical foundations of human decision-making in agent-based land use models-a review. *Environ Model Softw*. (2017) 87:39–48. doi: 10.1016/j.envsoft.2016.10.008

15. Schwarz N, Dressler G, Frank K, Jager W, Janssen M, Müller B, et al. Formalising theories of human decision-making for agent-based modelling of social-ecological systems: practical lessons learned and ways forward. *Socio Environ Syst Model*. (2020) 2:16340. doi: 10.18174/sesmo.2020a16340

16. Schlüter M, Baeza A, Dressler G, Frank K, Groeneveld J, Jager W, et al. A framework for mapping and comparing behavioural theories in models of social-ecological systems. *Ecol Econ*. (2017) 131:21–35. doi: 10.1016/j.ecolecon.2016.08.008

17. Taillandier F, Di Maiolo P, Taillandier P, Jacquenod C, Rauscher-Lauranceau L, Mehdizadeh R. An agent-based model to simulate inhabitants' behavior during a flood event. *Int J Disaster Risk Reduct*. (2021) 64:102503. doi: 10.1016/j.ijdrr.2021.102503

18. Vo DA. An operational architecture to handle multiple levels of representation in agent-based models. *Paris 6;*. (2012).

19. Gotts NM, van Voorn GA, Polhill JG, de Jong E, Edmonds B, Hofstede GJ, et al. Agent-based modelling of socio-ecological systems: models, projects and ontologies. *Ecol Complexity*. (2019) 40:100728. doi: 10.1016/j.ecocom.2018.07.007

20. Vo DA, Drogoul A, Zucker JD. An operational meta-model for handling multiple scales in agent-based simulations. In: *2012 IEEE RIVF International Conference on Computing & Communication Technologies, Research, Innovation, and Vision for the Future*. Ho Chi Minh City: IEEE (2012). p. 1–6.

21. Morvan G, Jolly D. Multi-level agent-based modeling with the Influence Reaction principle. *arXiv preprint arXiv:12040634*. (2012). doi: 10.48550/arXiv.1204.0634

22. Mathieu P, Morvan G, Picault S. Multi-level agent-based simulations: four design patterns. *Simulat Model Pract Theory*. (2018) 83:51–64. doi: 10.1016/j.simpat.2017.12.015

23. Henderson LF. On the fluid mechanics of human crowd motion. *Transport Res*. (1974) 8:509–15. doi: 10.1016/0041-1647(74)90027-6

24. Parr T, Sajid N, Friston KJ. Modules or mean-fields? *Entropy*. (2020) 22:552. doi: 10.3390/e22050552

25. Abouaissa H, JOLLY D, BENASSER A, et al. Macro-micro simulation of traffic flow. *IFAC Proc Vol*. (2006) 39:351–6. doi: 10.3182/20060517-3-FR-2903.00189

26. Crociani L, Vizzari G, Yanagisawa D, Nishinari K, Bandini S. Route choice in pedestrian simulation: design and evaluation of a model based on empirical observations. *Intell Artif*. (2016) 10:163–82. doi: 10.3233/IA-160102

27. Kiselev AV, Karbovskii VA, Kovalchuk SV. Agent-based modelling using ensemble approach with spatial and temporal composition. *Procedia Comput Sci*. (2016) 80:530–41. doi: 10.1016/j.procs.2016.05.333

28. Kadanoff LP. More is the same; phase transitions and mean field theories. *J Stat Phys*. (2009) 137:777–97. doi: 10.1007/s10955-009-9814-1

29. Auger P, Parra R, Poggiale JC, Sánchez E, Nguyen-Huu T. Aggregation of variables and applications to population dynamics. In: *Structured Population Models in Biology and Epidemiology*. Berlin: Springer (2008). p. 209–63.

30. Nguyen ND. *Coupling Equation-Based and Individual-based Models in The Study of Complex Systems: A Case Study in Theoritical Population Ecology*. Paris (2010).

31. Nguyen ND, Drogoul A, Auger P. Methodological steps and issues when deriving individual based-models from equation-based models: a case study in population dynamics. In: *Pacific Rim International Conference on Multi-Agents*. Berlin: Springer (2008). p. 295–306.

32. Nguyen DN, Taillandier P, Drogoul A, Auger P. Infering equation-based models from individual-based models. In: *Proceeding in PRIMA Conference*. Berlin: Springer (2010). p. 183–90.

33. Nguyen ND, Phan THD, Nguyen TNA, Drogoul A, Zucker JD. Disk graph-based model: a graph theoretical approach for linking agent-based models and dynamical systems. In: *2010 IEEE RIVF International Conference on Computing & Communication Technologies, Research, Innovation, and Vision for the Future (RIVF)*. Hanoi: IEEE (2010). p. 1–4.

34. Thuy NP. *Marker Assisted Selected for Enhanced Zinc Content in Rice: Nguyen Phuong Thuy*. University of Agricultural Sciences, GKVK. (2018).

35. Lèye B, Nguyen DN, Monga O, Garnier P, Nunan N. Simulating biological dynamics using partial differential equations: application to decomposition of organic matter in 3D soil structure. *Viet J Math*. (2015) 43:801–17. doi: 10.1007/s10013-015-0159-6

36. Nguyen-Trong K, Nguyen-Thi-Ngoc A, Nguyen DN, Dinh-Thi-Hai V. Optimization of municipal solid waste transportation by integrating GIS analysis, equation-based, and agent-based model. *Waste Manag*. (2017) 59:14–22. doi: 10.1016/j.wasman.2016.10.048

37. Rijpma JA. Complexity, tight-coupling and reliability: connecting normal accidents theory and high reliability theory. *J Contingencies Crisis Manag*. (1997) 5:15–23. doi: 10.1111/1468-5973.00033

38. Tchappi IH, Galland S, Kamla VC, Kamgang JC. A brief review of holonic multi-agent models for traffic and transportation systems. *Procedia Comput Sci*. (2018) 134:137–44. doi: 10.1016/j.procs.2018.07.154

39. Pike T. Multi-level mesa. *arXiv preprint arXiv:190408315*. (2019). doi: 10.48550/arXiv.1904.08315

40. Marcenac P, Giroux S. Geamas: a generic architecture for agent-oriented simulations of complex processes. *Appl Intell*. (1998) 8:247–67. doi: 10.1023/A:1008220501261

41. Combes M, Grigné C, Husson L, Conrad C, Le Yaouanq S, Parenthoën M, et al. Multiagent simulation of evolutive plate tectonics applied to the thermal evolution of the Earth. *Geochem Geophys Geosyst*. (2012) 13:4014. doi: 10.1029/2011GC004014

42. Steiniger A, Krüger F, Uhrmacher AM. Modeling agents and their environment in multi-level-DEVS. In: *Proceedings of the 2012 Winter Simulation Conference (WSC)*. Berlin: IEEE (2012). p. 1–12.

43. Concepcion AI, Zeigler BP. DEVS formalism: a framework for hierarchical model development. *IEEE Trans Softw Eng*. (1988) 14:228–41. doi: 10.1109/32.4640

44. Himmelspach J, Uhrmacher AM. The JAMES II framework for modeling and simulation. In: *2009 International Workshop on High Performance Computational Systems Biology*. Trento: IEEE (2009). p. 101–2.

45. Cossentino M, Gaud N, Galland S, Hilaire V, Koukam A. A holonic metamodel for agent-oriented analysis and design. In: *International Conference on Industrial Applications of Holonic and Multi-Agent Systems*. Berlin: Springer (2007). p. 237–46.

46. Lin Y, Hilaire V, Gaud N, Koukam A. K-crio: an ontology for organizations involved in product design. In: *International Conference on Digital Information and Communication Technology and Its Applications*. Berlin: Springer (2011). p. 362–76.

47. Missaoui E, Mazigh B, Bhiri S, Hilaire V. Ncrio: A normative holonic metamodel for multi-agent systems. In: *International Conference on Hybrid Artificial Intelligence Systems*. Berlin: Springer (2017). p. 638–49.

48. Hjorth A, Head B, Brady C, Wilensky U. Levelspace: a netlogo extension for multi-level agent-based modeling. *J Artif Societies Soc Simulat*. (2020) 23:4130. doi: 10.18564/jasss.4130

49. Tisue S, Wilensky U. Netlogo: a simple environment for modeling complexity. In: *International Conference on Complex Systems. Vol. 21*. Boston, MA (2004). p. 16–21.

50. Taillandier P, Gaudou B, Grignard A, Huynh QN, Marilleau N, Caillou P, et al. Building, composing and experimenting complex spatial models with the GAMA platform. *Geoinformatica*. (2019) 23:299–322. doi: 10.1007/s10707-018-00339-6

51. Reynolds CW. Flocks, herds and schools: a distributed behavioral model. In: *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques*. Berlin: Springer (1987) p. 25–34.

52. Vo DA, Drogoul A, Zucker JD, Ho TV. A modelling language to represent and specify emerging structures in agent-based model. In: *International Conference*

on Principles and Practice of Multi-Agent Systems. Berlin: Springer (2010). p. 212–27.

53. Lepagnot J, Hutzler G. A multiscale agent-based model for the simulation of avascular tumour growth. *J Biol Phys Chem*. (2009) 9:17–25. doi: 10.4024/11LE09A.jbpc.09.01

54. Bosmans S, Bogaerts T, Casteels W, Mercelis S, Denil J, Hellinckx P. Adaptivity in multi-level traffic simulation using experimental frames. *Simulat Model Pract Theory*. (2022) 114:102395. doi: 10.1016/j.simpat.2021.102395

55. Grignard A, Macià N, Alonso Pastor L, Noyman A, Zhang Y, Larson K. Cityscope andorra: a multi-level interactive and tangible agent-based visualization. In: *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*. Berlin: Springer (2018). p. 1939–40.

56. Servat D, Perrier E, Treuil JP, Drogoul A. When agents emerge from agents: introducing multi-scale viewpoints in multi-agent simulations. In: *International Workshop on Multi-Agent Systems and Agent-Based Simulation*. Berlin: Springer (1998). p. 183–98.

57. Chapuis K, Taillandier P, Gaudou B, Brugière A, Drogoul A, Araldi A, et al. Using the COMOKIT model to study the impact of the morpho-functional organization of cities on the spread of COVID-19. In: *6th International Workshop on Agent-Based Modelling of Urban Systems (ABMUS@ AAMAS 2021)*. Berlin: Springer (2021). p. 61–5.

58. Kaye D. *Loosely Coupled: The Missing Pieces of Web Services*. RDS Strategies LLC (2003).

59. Chapuis K, Elwaqoudi TA, Brugière A, Daudé E, Drogoul A, Gaudou B, et al. An agent-based co-modeling approach to simulate the evacuation of a population in the context of a realistic flooding event: a case study in hanoi (Vietnam). In: *International Workshop on Complex Systems Modelling & Simulation*. Berlin: Springer (2019). p. 79–108.

60. van Toll W, Braga C, Solenthaler B, Pettré J. Extreme-density crowd simulation: combining agents with smoothed particle hydrodynamics. In: *Motion, Interaction and Games*. Berlin: Springer (2020). p. 1–10.

61. Lu Y, Basak K, Carrion C, Loganathan H, Adnan M, Pereira FC, et al. SimMobility mid-term simulator: a state of the art integrated agent based demand and supply model. In: *Transportation Research Board 94th Annual MeetingTransportation Research Board*. Berlin: Springer (2015). p. 15–3937.

62. Soares G, Kokkinogenis Z, Macedo JL, Rossetti RJ. Agent-based traffic simulation using sumo and jade: an integrated platform for artificial transportation systems. In: *Simulation of Urban MObility User Conference*. Berlin: Springer (2013). p. 44–61.

63. Krajzewicz D, Hertkorn G, Rössel C, Wagner P. SUMO (Simulation of Urban MObility)-an open-source traffic simulation. In: *Proceedings of the 4th middle East Symposium on Simulation and Modelling (MESM20002)*. Berlin: Springer (2002). p. 183–7.

64. Bellifemine F, Bergenti F, Caire G, Poggi A. JADE–a java agent development framework. In: *Multi-Agent Programming*. Berlin: Springer (2005). p. 125–47.

65. David D, Payet D, Courdier R. Réification de zones urbaines émergentes dans un modèle simulant l'évolution de la population à La Réunion. In: *JFSMA*. Berlin: Springer (2011). p. 63–72.

66. Corchado JM, Mata A, Rodriguez S. Osm: a multi-agent system for modeling and monitoring the evolution of oil slicks in open oceans. In: *Advanced Agent-Based Environmental Management Systems*. Berlin: Springer (2009). p. 91–117.

67. Huynh QN. *CoModels, Engineering Dynamic Compositions of Coupled Models to Support the Simulation of Complex Systems. Université Pierre et Marie Curie-Paris VI*. Berlin: Springer (2016).

68. Butner JD, Cristini V, Wang Z. Development of a three dimensional, multiscale agent-based model of ductal carcinoma in situ. In: *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. Jeju: IEEE (2017). p. 86–89.

69. Wang Z, Maini PK. Editorial special section on multiscale cancer modeling. *IEEE Trans Biomed Eng*. (2017) 64:501–503. doi: 10.1109/TBME.2017.2655439

70. Alqurashi R, Altman T. Hierarchical agent-based modeling for improved traffic routing. *Appl Sci*. (2019) 9:4376. doi: 10.3390/app9204376

71. Rahman MM, Nguyen R, Lu L. Multi-level impacts of climate change and supply disruption events on a potato supply chain: an agent-based modeling approach. *Agric Syst*. (2022) 201:103469. doi: 10.1016/j.agsy.2022.103469

72. Yang L, van Dam KH, Anvari B, Zhang L. Multi-level agent-based simulation for supporting transit-oriented development in Beijing. In: *International Workshop on Agent-Based Modelling of Urban Systems (ABMUS)*. Berlin: Springer (2021). p. 17.

73. Ferber J, Gutknecht O, Michel F. From agents to organizations: an organizational view of multi-agent systems. In: *International Workshop on Agent-Oriented Software Engineering*. Berlin: Springer (2003). p. 214–30.

74. Ferber J. *Multi-Agent Systems: An Introduction to Distributed Artificial Intelligence*. Addison-Wesley Longman Publishing Co., Inc. (1999).

75. Ferber J, Gutknecht O. A meta-model for the analysis and design of organizations in multi-agent systems. In: *Proceedings International Conference on Multi Agent Systems (Cat. No. 98EX160)*. Paris: IEEE (1998). p. 128–35.

76. Ferber J, Stratulat T, Tranier J. Towards an integral approach of organizations in multi-agent systems. In: *Handbook of Research on Multi-Agent Systems: Semantics and Dynamics of Organizational Models*. IGI Global (2009). p. 51–75.

77. Benoudina L, Redjimi M. Multi-agent system based approach for industrial process simulation. *IEEE Internet Things J*. (2021) 54:209–17. doi: 10.18280/jesa.540202

78. Gutknecht O, Ferber J. MadKit: organizing heterogeneity with groups in a platform for multiple multi-agent systems. *Rapport Interne LIRMM*. (1997) 97188:1997.

79. Gutknecht O, Ferber J. Madkit: a generic multi-agent platform. In: *Proceedings of the Fourth International Conference on Autonomous Agents*. (2000). p. 78–9.

80. Abrami G, Lardon S, Barreteau O, Cernesson F. ORIGAMI, une méthode organisation centrée de modélisation multi-agent de systèmes complexes. *Revue Internationale de Géomatique*. (2006) 16: 211. doi: 10.3166/rig.16.211-232

81. Barreteau O, Garin P, Dumontier A, Abrami G, Cernesson F. Agent-based facilitation of water allocation: case study in the Drome River Valley. *Group Dec Negot*. (2003) 12:441–61. doi: 10.1023/B:GRUP.0000003743.65698.78

82. Ferber J, Michel F, Báez J. AGRE: integrating environments with organizations. In: *International Workshop on Environments for Multi-Agent Systems*. Berlin: Springer (2004). p. 48–56.

83. Michel F. The IRM4S model: the influence/reaction principle for multiagent based simulation. In: *Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems*. Berlin: Springer (2007). p. 1–3.

84. Ferber J, Müller JP. Influences and reaction: a model of situated multiagent systems. In: *Proceedings of Second International Conference on Multi-Agent Systems (ICMAS-96)*. (1996). p. 72–9.

85. Gechter F, Contet JM, Galland S, Lamotte O, Koukam A. Virtual intelligent vehicle urban simulator: application to vehicle platoon evaluation. *Simulat Model Pract Theory*. (2012) 24:103–14. doi: 10.1016/j.simpat.2012.02.001

86. Soyez JB, Morvan G, Dupont D, Merzouki R. A methodology to engineer and validate dynamic multi-level multi-agent based simulations. In: *International Workshop on Multi-Agent Systems and Agent-Based Simulation*. Berlin: Springer (2012). p. 130–42.

frontiers | Frontiers in Applied Mathematics and Statistics

TYPE Original Research
PUBLISHED 05 December 2022
DOI 10.3389/fams.2022.1045241

OPEN ACCESS

EDITED BY
Junpyo Park,
Kyung Hee University-Global Campus, South Korea

REVIEWED BY
Halim Zeghdoudi,
University of Annaba, Algeria
Mohd Tahir Ismail,
Universiti Sains Malaysia (USM), Malaysia

*CORRESPONDENCE
Dodi Devianto
ddevianto@sci.unand.ac.id

SPECIALTY SECTION
This article was submitted to Mathematics of Computation and Data Science, a section of the journal Frontiers in Applied Mathematics and Statistics

RECEIVED 15 September 2022
ACCEPTED 07 November 2022
PUBLISHED 05 December 2022

CITATION
Devianto D, Ramadani K, Maiyastri, Asdi Y and Yollanda M (2022) The hybrid model of autoregressive integrated moving average and fuzzy time series Markov chain on long-memory data.
Front. Appl. Math. Stat. 8:1045241.
doi: 10.3389/fams.2022.1045241

The hybrid model of autoregressive integrated moving average and fuzzy time series Markov chain on long-memory data

Dodi Devianto*, Kiki Ramadani, Maiyastri, Yudiantri Asdi and Mutia Yollanda

Department of Mathematics and Data Science, Andalas University, Padang, Indonesia

Introduction: The price of crude oil as an essential commodity in the world economy shows a pattern and identifies the component factors that influence it in the short and long term. The long pattern of the price movement of crude oil is identified by a fractionally time series model where the accuracy can still be improved by making a hybrid residual model using a fuzzy time series approach.

Methods: Time series data containing long-memory elements can be modified into a stationary model through the autoregressive fractional integrated moving average (ARFIMA). This fractional model can provide better accuracy on long-memory data than the classic autoregressive integrated moving average (ARIMA) model. The long-memory data are indicated by a high level of fluctuation and the autocorrelation value between lags that decreases slowly. However, a more accurate model is proposed as a hybridization time series model with fuzzy time series Markov chain (FTSMC).

Results: The time series data collected from the monthly period of West Texas Intermediate (WTI) oil price as the standard for world oil prices for the 2003–2021 time period. The data of WTI oil price has a long-memory data pattern to be modeled fractionally, and subsequently their hybrids. The times series model of crude oil price is obtained as the new target model of hybrid ARIMA and ARFIMA with FTSMC, denoted as ARIMA-FTSMC and ARFIMA-FTSMC, respectively.

Discussion: The accuracy model measured by MAE, RMSE, and MAPE shows that the hybrid model of ARIMA-FTSMC has better performance than ARIMA and ARFIMA, but the hybrid model of ARFIMA-FTSMC provides the best accuracy compared to all models. The superiority of the hybrid time series model of ARFIMA-FTSMC on long-memory data provides an opportunity for the hybrid model as the best and more precise forecasting method.

KEYWORDS

autoregressive integrated moving average, autoregressive fractionally integrated moving average, fuzzy time series Markov, hybrid time series model, model accuracy

1. Introduction

Crude oil is a commodity that plays an essential role in the economy. Fluctuations in crude oil prices can impact the health of the world economy, especially in crude oil-producing countries. This condition occurs because countries that import crude oil are more susceptible to price changes with transactions carried out in US dollars (USD). One of the crude oil types used as a benchmark in determining oil prices is West Texas Intermediate (WTI). WTI oil, according to the U.S. Energy Information Administration (EIA) website, is included as a type of petroleum that is low in density. WTI oil contains about 0.24% sulfur and a gravity of 39.6°. This oil is usually considered of good quality for processing into gasoline. Therefore, this high-quality WTI oil is used as the standard price for world oil.

The fluctuation of crude oil prices shows a pattern and identifies the component factors that influence it in the short and long term. The classical time series model of autoregressive integrated moving average (ARIMA) is often used to produce accurate short memory modeling [1]. In nonlinear time series data, the accuracy of the ARIMA model in forecasting is quite good than the recurrence quantification analysis (RQA) predictive model [2]. However, the heteroscedasticity in the ARIMA model can be corrected by using a variance model, that is the GARCH model [3] or the GARCH exponential [4] or the mixed memory MMGARCH [5].

Oil price movements are very volatile and tend to be affected for a long time. Thus, there are some cases of time-series data showing a long-memory pattern. In brain stimulation, for instance, this volatility problem as the dynamical pattern is extended through stimulation for long-term targeting and control of post-stimulation [6]. This long-memory case is characterized by decaying correlation slowly over time at infinity [7]. As a result, the value of the differentiating coefficient in the form of an integer number cannot provide an accurate estimate in the ARIMA model, so the fractional differentiating number of the ARFIMA model is used instead [8]. The method for estimating the value of the fractional differentiating coefficient in the ARFIMA model that is often used is the Geweke and Porter-Hudak (GPH) method to estimate the differentiating coefficient parameter directly without knowing the value of the autoregressive (AR) or the moving average (MA) order [9]. In addition, the Whittle estimator is provided for obtaining the parameter estimates for the stationary time series model of ARMA and ARFIMA, which is the basis of the Jackknifed empirical likelihood inference [10].

Classical time series data modeling require many assumptions. Thus, many have developed time-series data modeling using a fuzzy logic approach, the fuzzy time series (FTS) method. The simple arithmetic operations on time series data were applied to develop the early period of FTS, while a fuzzy logic was proposed by using weighting and adaptive modeling [11]. Finally, the FTS model has been applied in forecasting oil production and consumption [12]. Further developments in 2012, the fuzzy time series Markov chain (FTSMC) was first proposed as a new concept to analyze the accuracy of currency exchange rate predictions. This FTSMC method gives quite good results compared to the FTS method [13]. Furthermore, the forecasting method of FTSMC was also carried out on gold prices as the investment information [14]. In addition, three FTS methods, namely FTS Chen, FTS Segmented Chen, and FTSMC are compared to forecast bitcoin prices [2]. Based on the mean absolute percentage error (MAPE) accuracy value, the FTSMC method still gave better results. These results confirmed that the FTSMC method provides a pretty good modeling accuracy than other FTS methods.

Time series data modeling with FTS has many parameters because the model changes in each iteration process. A simple structure has been proposed to overcome this model change. Parsimony FTS is used in forecasting prices for liquid bulk cargo carriers and secondhand ships [15]. In advanced financial cases, the time series data of the Indonesian Composite index price was analyzed by using the nonlinear artificial intelligence method, which produced the best accuracy based on the MAPE value [16]. Meanwhile, a development of the hybrid ARIMA model and a wavelet-based artificial intelligence model gave more accurate results than the ARIMA model or artificial intelligence model [17]. However, ARIMA, ARFIMA, and FTS models have something in common: using past values to produce modeling in the future period. The difference is in the residual assumption test that must be met using the classical ARIMA and ARFIMA models. These assumptions are not a concern for modeling with the FTS method. Meanwhile, the residual of model ARIMA and ARFIMA can be adjusted in their value and residual assumptions by combining the model of ARIMA and ARFIMA with the numerical data processing of the fuzzy time series Markov chain. This study proposes the best modeling for West Texas Intermediate (WTI) crude oil prices using the hybrid model of ARIMA-FTSMC and the hybrid model of ARFIMA-FTSMC hybrid models, compared to the classical ARIMA and ARFIMA models based on the level of accuracy measured using mean absolute percentage error (MAPE), mean absolute error (MAE), and root mean square error (RMSE).

2. Materials and methods

This study uses available data from West Texas Intermediate Oil (WTI) on the U.S. Energy Information Administration (EIA) website with the monthly period during the 2003–2021 time period, consisting of 18 years selected as many as 217 data. The time series data used are long-memory which will be modeled into classical time series models and hybrid models. Furthermore, in this section, theories and processes related to the formation of target models will be explained, namely the model of ARIMA, the model of ARFIMA, the hybrid model of ARIMA-FTSMC, and the hybrid model of ARFIMA-FTSMC.

2.1. Autoregressive integrated moving average

A time series $\{X_t\}$ has the properties of white noise if a sequence of uncorrelated random variables with a specific distribution is identified by constant mean, usually assumed to be 0, a constant variance $Var(X_t) = \sigma^2$ and $Cov(X_{t+h}, X_t) = 0$ for $k \neq 0$. In time series analysis, there are some time series models such as ARIMA which is a combined two models between autoregressive (AR) and moving average (MA) after differencing. The common form of the ARIMA model is expressed as follows:

$$\phi_p(B)(1-B)^d X_t = \theta_q(B)\varepsilon_t \tag{1}$$

with

$$\begin{aligned}\phi_p(B) &= (1-\phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)\\ \theta_q(B) &= (1-\theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)\end{aligned}$$

where $\phi_p(B)$ is the autoregressive components, $\theta_q(B)$ is the moving average components, B is the operator of backward shift, and $(1-B)^d X_t$ is stationary of time series in d-order differencing. This process is denoted by ARIMA(p,d,q).

For detecting the stationarity of data, graph analysis can not be proposed to determine whether the time series data are already stationary, but it helps to know how the pattern of the data. The basic properties are still needed to determine the next decision. If the data have a constant mean and variance, the data are already stationary. If the variance of the data is non-stationary, it can be solved by using the power transformation, namely the Box–Cox transformation. Let $T(X_t)$ be the transformation function of X_t. The following formula is used to stabilize the variance

$$T(X_t) = \frac{X_t^{\lambda} - 1}{\lambda} \tag{2}$$

for $\lambda \neq 0$ and λ called transformation parameter. After the data are stationary in variance, it is followed by testing the stationary in the mean by using augmented Dickey–Fuller (ADF) test. The random walk equation with drift for the differenced-lag model is regressed to be:

$$\nabla X_t = \mu + \delta X_{t-1} + \sum_{i=1}^{k} \phi_i \nabla X_{t-i} + e_t \tag{3}$$

for $\nabla X_t = X_t - X_{t-1}$, k is the number of lags, δ is the slope coefficient, μ is a drift parameter, ϕ_i is the parameter of random walk equation, and e_t is the white noise error term. The test statistic is used as follows:

$$ADF = \frac{\hat{\delta}}{SE(\hat{\delta})} \tag{4}$$

for $\hat{\delta}$ as the estimated δ which is obtained by using ordinary least squares and $SE(\hat{\delta})$ as the standard error of δ. The initial or null hypothesis of $\delta = 0$ means that the stationarity has not been fulfilled. The conclusion of the ADF hypothesis test is rejecting the null hypothesis if the ADF value is less than the test statistics.

The model of autoregressive integrated moving average (ARIMA) can be built by using the following steps.

1. Checking the stationarity of the data against the variance with Box–Cox transformation and the stationarity of the mean value by differencing and augmented Dickey–Fuller (ADF) test.
2. Identifying possible ARIMA models by determining orders based on lag correlation analysis on ACF and Partial ACF (PACF) plots.
3. Estimating significance parameter on ARIMA model with significant level $\alpha = 5\%$.
4. Choosing the best ARIMA model that has the smallest value of Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values using the formulation below

$$AIC = n\ ln(\hat{\sigma_\varepsilon}^2) + 2k \tag{5}$$

$$BIC = n\ ln(\hat{\sigma_\varepsilon}^2) + kln(n) \tag{6}$$

for n is the number of observations, $\hat{\sigma}_\varepsilon^2$ is the maximum likelihood estimator of σ_ε^2, and k is the number of parameters estimated.

5. Testing the ARIMA model residual assumptions using non-autocorrelation, homoscedasticity, and normality tests.

   a. Non-autocorrelation test using QLjung-Box test with the equation:

$$Q_{LB} = n(n+2)\sum_{i=1}^{k} \frac{\rho_i^2}{n-i} \tag{7}$$

   where k is the number of lag, n is the number of observations, and ρ_i^2 is the autocorrelation of the i-th residual for $i = 1, 2, ..., k$. The value of Q_{LB} follows χ^2 distribution with a degree of freedom $k - p - q$ where p and q are the order in the ARIMA model. If $Q_{LB} < \chi_\alpha^2(k-p-q)$, then the residual in the model is non-autocorrelation.

   b. The homoscedasticity test was carried out using the Lagrange multiplier test with the test statistic,

$$LM = nR^2 \tag{8}$$

   where n is the number of observations and R^2 is the coefficient of determination of the quadratic residual regression model. The LM value follows χ^2 distribution with a degree of freedom q which is one order in the

ARIMA model. If $LM < \chi^2_\alpha(q)$, then the residual in the model is homoscedasticity.

c. The normality test was carried out using the Jarque–Bera hypothesis with the test statistic

$$JB = \frac{n}{6}\left(S^2 + \frac{(K-3)^2}{4}\right) \tag{9}$$

where n is the number of observations,

$$S = \frac{\frac{1}{n}(\varepsilon_i - \bar{\varepsilon})^3}{\left(\frac{1}{n}\sum_{i=1}^{n}(\varepsilon_i - \bar{\varepsilon})^2\right)^{\frac{3}{2}}} \quad K = \frac{\frac{1}{n}(\varepsilon_i - \bar{\varepsilon})^4}{\left(\frac{1}{n}\sum_{i=1}^{n}(\varepsilon_i - \bar{\varepsilon})^2\right)^2},$$

and $\bar{\varepsilon} = \frac{\sum_{i=1}^{n}\varepsilon_i}{n}$

The JB value follows χ^2 distribution with a degree of freedom of 2. If the $JB < \chi^2_\alpha(2)$, then the residual in the model is normally distributed.

2.2. Autoregressive fractional integrated moving average

Autoregressive fractionally integrated moving average (ARFIMA) was first introduced by Grager and Joyeux in 1980 which is the development of the autoregressive integrated moving average (ARIMA) model to model long-term data. Long memory is a stationary time series that has a long-term dependence between observations with periods that are far apart but still have a high correlation. This can also be seen from the autocorrelation function that decays slowly over a long period. The ARFIMA(p, d, q) model, where p and q are non-negative integers and d is a real number in the interval $0 < d < 0.5$. The general model ARFIMA(p, d, q) can be shown as follows Wei WWS [1]:

$$\phi_p(B)(1-B)^d X_t = \theta_q(B)\varepsilon_t \tag{10}$$

where $\phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p)$, $\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q)$, B as backshift operator, d as the differencing in real number, $(1-B)^d X_t$ as the time series that has been already at differencing, θ_i as the parameter of MA for i, $i = 1, 2, ..., q$, ϕ_i as the parameter of AR for i, $i, 1, 2, ..., p$, and $\varepsilon_t \sim WN(0, \sigma^2)$.

The model building of the autoregressive fractional integrated moving average (ARFIMA) can be determined by using the following steps.

1. Checking the stationary of the data against variance. If the data are not stationary concerning variance, a transformation is carried out using the Box–Cox transformation method to obtain a rounded value (λ).
2. Checking the stationary of the data in the mean using augmented Dickey–Fuller (ADF). If the data are not stationary concerning the mean, then a differencing is required.
3. Estimating differencing parameters using the GPH method with the following formula:

$$\widehat{d}_{GPH} = \frac{\sum_{j=1}^{m}(x_j - \bar{x})(y_j - \bar{y})}{\sum_{j=1}^{m}(x_j - \bar{x})^2}. \tag{11}$$

where $y_j = lnI(\lambda_j)$ and $x_j = -ln(2sin(\frac{\lambda_j}{2}))^2$. The $I(\lambda_j)$ function is a periodogram with a frequency of Fourier $\lambda_j = \frac{2\pi j}{T}$, $j = 1, 2, \ldots, m$, and T is the number of observation data, while m is the limit of the number of Fourier frequencies.

4. Differentiating data that have been transformed using the value of $\hat{d}_{GPH}$.
5. Identifying the ARFIMA model by determining the combination of the ARFIMA model by making ACF and PACF plots from the differencing of data. The ACF plot shows the order of $MA(q)$, and the plot of the PACF shows the order of $AR(p)$.
6. Parameter estimation and significance test of the ARFIMA model have been declared significant if the model parameter's probability value is smaller than $\alpha = 5\%$.
7. Choosing the best ARFIMA model is the model that has the smallest AIC and BIC values.
8. Testing the best ARFIMA model residual assumptions, namely non-autocorrelation, heteroscedasticity, and normality tests.

2.3. Proposed model of hybrid ARIMA-FTSMC

The ARIMA-FTSMC hybrid model is a proposed model by combining the ARIMA model for initial data and the FTSMC model for residual data. The first step is to model the time series data with the best ARIMA model. In improving the ARIMA model, the residual ARIMA model is analyzed by using Geweke and Porter–Hudak (GPH) in Equation (11) to determine whether there still contains long-memory properties. The improvement of the residual ARIMA can apply the FTSMC model as the alternative method because the FTSMC model does not require stationary data and regression assumption. Finally, residual modeling results from the FTSMC model are entered into the ARIMA model to obtain the ARIMA-FTSMC hybrid modeling. The steps for modeling ARIMA residuals using the FTSMC model are as follows:

1. Definition of the union of sets U. In the first stage, determine the maximum value (D_{max}) and minimum value (D_{min}) from ARIMA data residual. Choose any positive

value for D_1 and D_2 so that they can be used in the formation of the union of sets U with the following conditions:

$$U = [D_{min} - D_1, D_{max} + D_2), \tag{12}$$

2. Interval formation and interval length. The union of sets U will be partitioned into parts with equal intervals (n). By using the Sturges formula as follows:

$$n = 1 + 3.322\ log\ N, \tag{13}$$

where N is the number of ARIMA residual data. Here is the formula for determining the length of the interval:

$$l = \frac{[(D_{max} + D_2) - (D_{min} - D_1)]}{n}, \tag{14}$$

and l is the length of the interval and n is the number of intervals. Each interval can be calculated as follows:

$$u_n = [D^* + (n-1)l, D^* + nl). \tag{15}$$

where $D^* = D_{min} - D_1$.

3. The definition of fuzzy sets for each interval. If the ARIMA residual data are in the interval u_i, then the fuzzy set of the data is denoted as A_i.
4. Determination of fuzzy logic relations (FLRs). If $F(t) = A_i$ and $F(t-1) = A_j$, then the relationship between $F(t)$ and $F(t-1)$ is called a fuzzy logical relationship (FLR). This relationship can be expressed by $A_i \rightarrow A_j$, where A_i is the left-hand side (LHS) and A_j is the right-hand side (RHS) from FLR.
5. Determination of fuzzy logic relations group (FLRG). If two FLRs have the same fuzzy set (LHS $A_i \rightarrow A_{j1}$, $A_i \rightarrow A_{j2}$), then it can be grouped into fuzzy logical relationship groups (FLRGs) $A_i \rightarrow A_{j1}, A_{j2}$.
6. Markov transition probability matrix formation.

$$P_{ij} = \frac{M_{ij}}{M_i},\ i, j = 1, 2, 3, ..., n, \tag{16}$$

Transition probability of state A_i to A_j in one step is P_{ij}. The number of data from state A_i is M_i. The transition time of state A_i to A_j in one step is M_{ij}. Transition probability matrix R using $n \times n$ dimension can be written as:

$$\mathbf{R} = \begin{bmatrix} P_{11} & P_{12} & \dots & P_{1n} \\ P_{21} & P_{22} & \dots & P_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ P_{n1} & P_{n2} & \dots & P_{nn} \end{bmatrix}. \tag{17}$$

7. Calculation of the initial modeling results based on the transition probability R matrix with the following rules:

a. If FLRG from A_i transitions to the empty set, $A_i \rightarrow \phi$, then the modeling results of $F(t)$ is m_i, where m_i is the median of u_i, with equation:

$$F_t = m_i. \tag{18}$$

b. If FLRG A_i transitions to, $A_i \rightarrow A_k$ and $P_{ik} = 1, j \neq k$, the modeling result of $F(t)$ is m_k as the median of u_k, with equation:

$$F_t = m_k P_{ik} = m_k. \tag{19}$$

c. If FLRG A_j makes a one-to-many transition, $A_i \rightarrow A_1, A_2, \dots, A_n, j = 1, 2, \dots, n$, and datasets X_{t-1} when $t-1$ is in state A_j, then the modeling results of $F(t)$ are as follow:

$$\begin{aligned} F_t = {} & m_1 P_{j1} + m_2 P_{j2} + \cdots + m_{j-1} P_{j(j-1)} + X_{t-1} P_{jj} \\ & + m_{j+1} P_{j(j+1)} + \cdots + m_{n-1} P_{j(n-1)} + m_n P_{jn}, \end{aligned} \tag{20}$$

where $m_1, m_2, \dots, m_{j-1}, m_{j+1}, \dots, m_n$ is the median of $u_1, u_2, \dots, u_{j-1}, u_{j+1}, \dots, u_n$. The m_j value was substituted by X_{t-1} in order to obtain information from the state A_j when $t-1$.

8. Calculation of modeling adjustment values that aim to correct modeling errors. This is due to the biased transition probability **R** matrix. The calculation of the modeling adjustment value is as follows:

a. If state A_i when $t-1$ as $F_{t-1} = A_i$, and t with equation ($1 \leq s \leq n-i$) there is a forward transition jump to state A_{i+s}, then D_t as the adjustment value is defined as $D_t = \left(\frac{l}{2}\right) s$ where $1 \leq s \leq n-i$, l is the length of the interval, and s is the number of forwarding transition displacement jumps.
b. If state A_i when $t-1$ as $F_{t-1} = A_i$, and when t as ($1 \leq v \leq i$) there is a transition jump backward transition to state A_{i-v}, then D_t as the adjustment value is defined as $D_t = -\left(\frac{l}{2}\right) v$, where $1 \leq v \leq i$, l is the length of the interval, and v is the number of jumps of the backward transition displacement.

9. Calculation of the final modeling results. The general form of the final modeling result is the form:

$$FM_t = F_t + D_t. \tag{21}$$

2.4. Proposed model of ARFIMA-FTSMC

In building the model for long-memory data, the ARFIMA model is already enough required. In this section, the proposed model of ARFIMA-FTSMC is built by combining the algorithms

of ARFIMA and FTSMC, where the FTSMC is applied for adjusting the residual of the ARFIMA model. Therefore, the new proposed model of ARFIMA-FTSMC will have better performance than the ARFIMA model.

2.5. Modeling accuracy and goodness-of-fit

Error calculation is a way to determine the level of accuracy of the model that has been obtained with the observation data. The use of modeling techniques with the smallest error rate is a good modeling technique. Methods to calculate the size of this error include the mean absolute percentage error (MAPE), mean absolute error (MAE), and root mean squared error (RMSE). In selecting the best model, the smallest value of MAPE, MAE, and RMSE is required. Calculation accuracy using X_t as observation data and $\hat{X}_t$ as modeling data, where the formula for determining the MAPE value is as follows:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|X_t - \hat{X}_t|}{X_t} \times 100\% \tag{22}$$

The MAPE accuracy criteria are as follows Zhang et al. [10]:

a. The modeling accuracy is perfect when the MAPE value is less than 10%
b. The modeling accuracy is good when the MAPE value is between 10 and 20%
c. The modeling accuracy is fair when the MAPE value is between 20 and 50%
d. Modeling accuracy is not accurate when the MAPE value is greater than 50%.

The determination of the RMSE and MAE values is given by the following formula:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(X_t - \hat{X}_t)^2} \tag{23}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|X_t - \hat{X}_t| \tag{24}$$

In measuring the goodness of fit, the most popular coefficient determination R^2 is required. This measure is obtained by computing the ratio of sums of squares of regression (SSR) to the sums of squares total (SST). The coefficient of determination R^2 has the proper range of 0 to 1 the low values indicating poor fit, while the large values indicate well fit. Let $\bar{y}$ as the mean of data set $y_i, i = 1, 2, \ldots, n$ so the R^2 can be defined as follows:

$$R^2 = \frac{\sum_{i=1}^{n}(y_i' - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{25}$$

The value of R^2 is defined as the proportion of variance in the response variable accounted for by knowledge of the predictor variable(s). R^2 is also simultaneously the squared correlation between observed values on y_i and predicted values y_i' based on the data processing [18].

3. Results and discussions

In this section, the model building is related to the theory so that the model processing of WTI oil prices using three

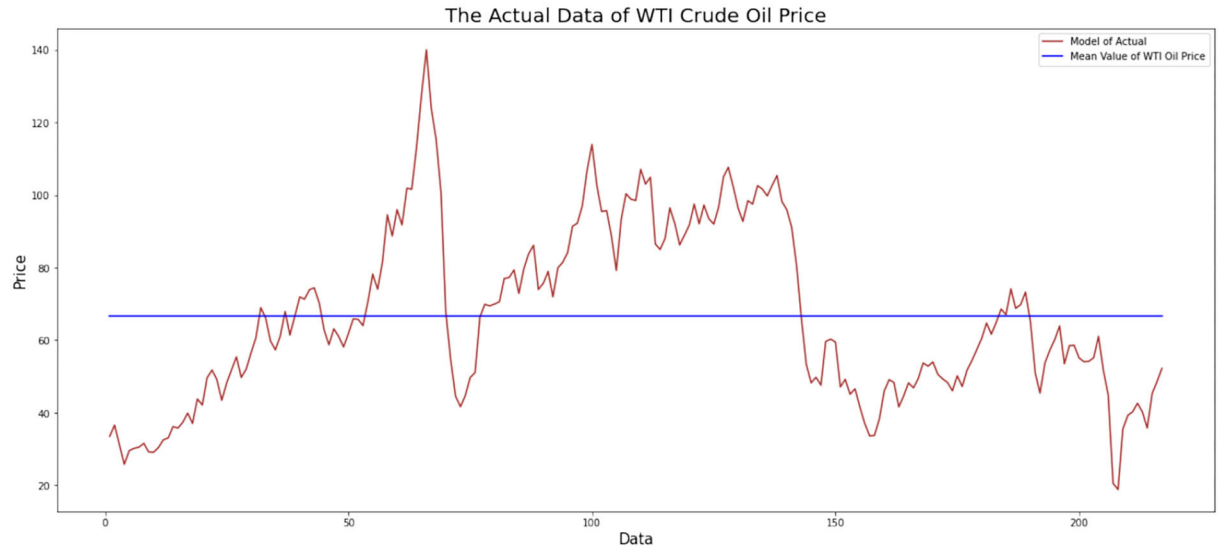


FIGURE 1
WTI oil price monthly data plot.

methods of the model of ARIMA, the model of ARFIMA, the hybrid model of ARIMA-FTSMC, and the hybrid model of ARFIMA-FTSMC will be explained in the following subsections.

3.1. The ARIMA model of WTI oil price

The first step in identifying a time series model is to identify it visually. This step can be figured by plotting the monthly data on WTI oil prices. This data plot aims to observe whether the data pattern has a trend, or seasonal component, and can also see the stationary of data. The plotting of monthly data on WTI oil prices from January 2003 to January 2021 is shown in Figure 1.

Figure 1 indicates that the monthly data on WTI oil prices has an uptrend and a downtrend at certain times. The data do not fluctuate around the mean, and the variance is not constant during observation. The data can be stationary to the variance and mean by performing a Box–Cox transformation and differentiating the data, respectively. The result of parameter Box–Cox, that is 0.3839, means that the data will be transformed as exponential with that parameter once to get the new data. At the same time, the uptrend and a downtrend fluctuation at certain times indicated that the ADF test is required. The result of the ADF test is shown in Table 1.

TABLE 1 Augmented Dickey–Fuller test.

Critical value	ADF test	
	Statistic value	*p*-value
1% : −3.4611		
5% : −2.8751	−2.8060	0.0574
10% : −2.5740		

It indicates that the mean of the data is not stationary yet. Furthermore, the exchange rate data have to make a difference. The plot of the transformed data on WTI oil prices is shown in Figure 2.

Based on Figure 2, it shows that the data have spread around the mean and variance. In other words, the data are stationary concerning the mean and variance. Next, identify the order of the ARIMA model by looking at the ACF and PACF plots as follows:

From Figure 3, it shows that the significant values of the ACF and PACF coefficients are the same at lag 1, so the ARIMA(p,d,q) model can be chosen with the order p and q equal to 1. And because differencing is carried out once, the order differencing is equal to 1. Thus, the possible models are ARIMA(1,1,0), ARIMA(1,1,1), and ARIMA(0,1,1). Next, the results of the estimation of the parameters for each model are shown in Table 2.

Then the parameter significance test is carried out by determining the probability value of the model parameters. The model is indicated to be significant if the probability value of the model is smaller than 0.05, which means that the ARIMA(1,1,0) and ARIMA(0,1,1) models are significant and ready to be applied. After testing the significance of the parameters, the next step is to choose the best model from the significant model by comparing the AIC and BIC values for each model which is presented in Table 3.

From the comparison of AIC and BIC values in Table 3 for the three models, it can be observed that the ARIMA(1,1,0) model has the smallest AIC and BIC values among other models. So it can be concluded that the best model is the ARIMA(1,1,0) model. The residual assumption test of the ARIMA (1,1,0) model is shown in Table 4.

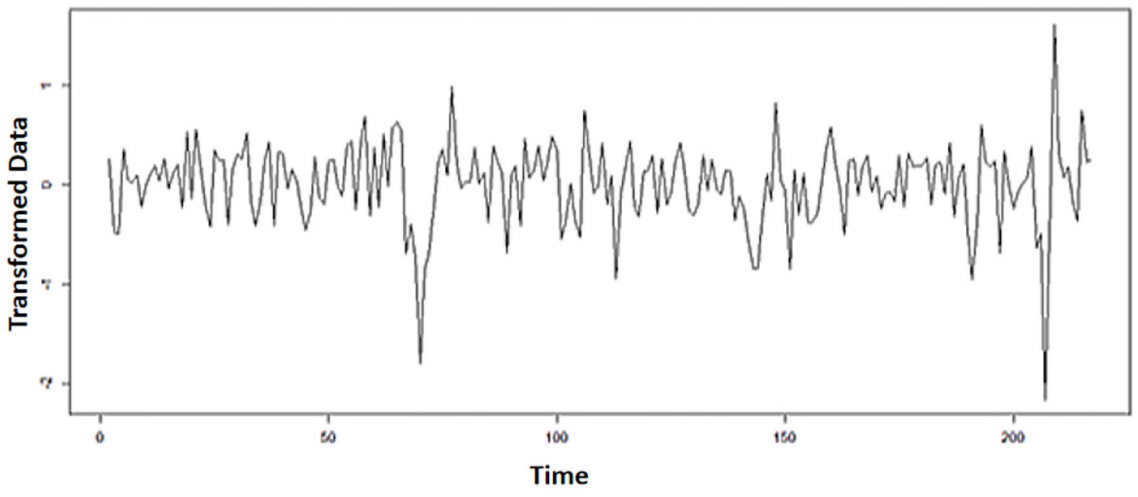


FIGURE 2
WTI oil data plot after transformation and differencing.

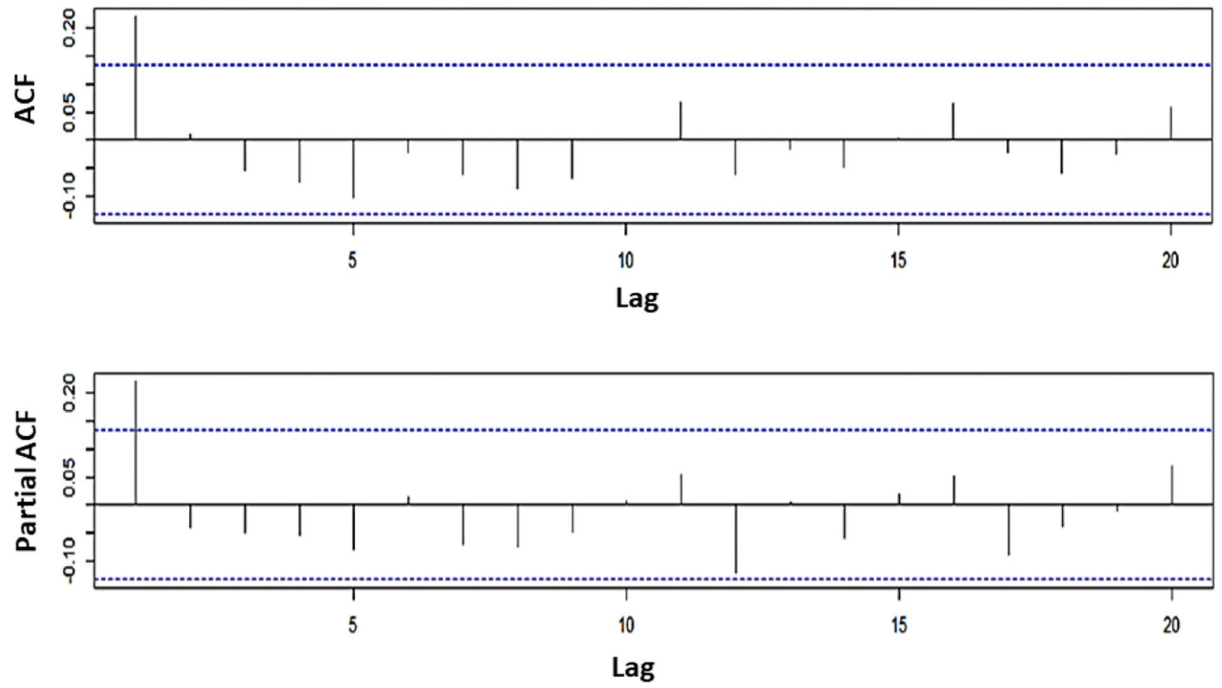


FIGURE 3
ACF and PACF plots of WTI oil data.

TABLE 2 Parameter estimation ARIMA(*p*,*d*,*q*) model with its probability value.

Model	ARIMA(1,1,0)	ARIMA(1,1,1)	ARIMA(0,1,1)
ϕ	0.2391	0.2722	-
	(0.0003*)	(0.2453)	
θ	-	−0.0351	0.2187
		(0.8841)	(0.0004*)

*Indicates the significant parameter.

TABLE 3 Comparison of AIC and BIC values ARIMA(*p*,*d*,*q*) model.

Model	ARIMA(1,1,0)	ARIMA(0,1,1)
AIC	1419.2900	1420.3750
BIC	1426.0410	1427.1250

TABLE 4 Residual assumption test ARIMA(1,1,0) model.

	Non-autocorrelation	Homoscedasticity	Normality
p-value	0.9508	0.3490	0.0882

Based on Table 4, it shows that the p-value of the non-autocorrelation and homoscedasticity test is p-value greater than 0.05. It means that there is no correlation between the data residuals, and the variance of the residuals is the same every time (homoscedasticity). Meanwhile, in the normality test, the *p*-value which is greater than 0.05 identifies that the residual data are normally distributed. So, the ARIMA(1,1,0) model is obtained as the best model that fulfills the residual assumptions.

3.2. The ARFIMA model of WTI oil price

The first step is to check that WTI oil data are stationary against variance. Next, consider the ACF plot of the stationary WTI oil data against the variance to see the long-memory pattern as shown in Figure 4.

Based on Figure 4, it shows that the data decrease slowly over time, which means that the data have a long-memory pattern. Furthermore, the estimation of differencing parameters is determined by using the GPH method in order to obtain the d_{GPH} value of d. Then, differencing the stationary data on the variance with a value of d_{GPH} has been obtained. Furthermore, the ARFIMA model was identified by looking at the ACF and PACF plots as follows:

Figure 5 indicates that the significant ACF coefficient value reaches lag 5, while the significant PACF coefficient value is at lag 1. Thus, there are 11 possible ARFIMA models. The next step

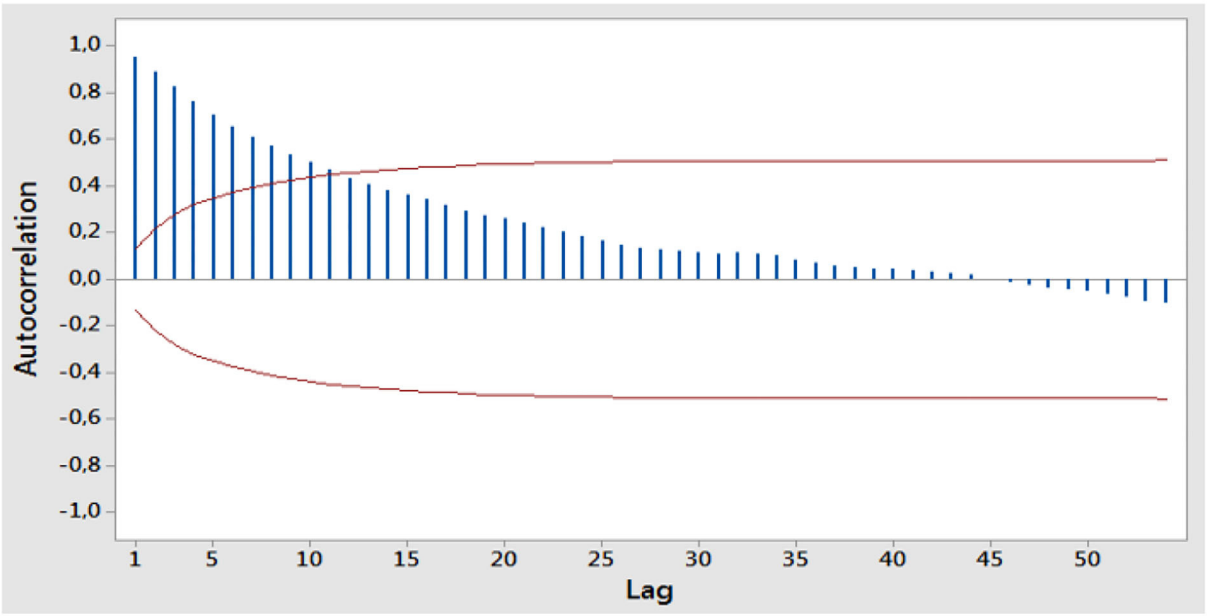


FIGURE 4
ACF plot of stationary WTI oil data against variance.

is to estimate the parameter and its probability value for each model, whose results are shown in Table 5.

Based on Table 5, the significant models have a smaller probability value than the significant level. There are ARFIMA(0, *d*, 1), ARFIMA(0, *d*, 2), ARFIMA(0, *d*, 3), ARFIMA(0, *d*, 4), and ARFIMA(1, *d*, 0) which are included as significant models and already used for forecasting new data. The next step is to choose the best model by comparing the AIC and BIC values for each model presented in Table 6.

From the comparison of AIC and BIC values in Table 7 for the five models, it can be observed that the ARFIMA(1,*d*.0) model has the smallest AIC and BIC values among other models. So it can be concluded that the best model is the ARFIMA(1,*d*,0) model. The following residual assumption test of the ARFIMA model (1,*d*.0) is shown in Table 8.

Based on Table 8, it describes that the non-autocorrelation test has a *p*-value greater than 0.05. It means that there is no correlation between the data residuals. Moreover, the normality test obtained a *p*-value greater than 0.05, which shows that the residuals are normally distributed. So the ARFIMA(1,*d*.0) model is obtained as the best model that fulfills the residual assumption.

3.3. The hybrid ARIMA-FTSMC model of WTI oil price

The hybrid ARIMA-FTSMC model is a combination of the ARIMA and FTSMC models. The first step is to build a model of time series data with the ARIMA model. The results show that the ARIMA(1,1,0) model is selected to be the best model. Thus, the residual data of the ARIMA(1,1,0) model, denoted by ε_t, are built using the FTSMC model. In the form of a negative number, the transformation is carried out so that the data become a non-negative number. The data transformation is used to determine the minimum data from the residual data. Then, the residual data are subtracted from the minimum data to become a non-negative number. Based on the residual data, the result shows the minimum residual data is $d_{min} = -29.2866$. Suppose the residual data at the time $t = 3$, which is $\varepsilon_3 = -6.2987$ so that the residual transformation data obtained is 22.9878. The first step in the modeling residual data using the FTSMC model is to determine the union of set U. Based on the residual transformation data, it is obtained that $D_{max} = 46.3287$ and $D_{min} = 0$. Then, the values are set as $D_1 = 0$ and $D_2 = 0.6713$ that is associated with D_{min} and D_{max} so that the lower and upper bound in the interval will always include the data to ignore the outlier data. The union of set U is defined as an interval below:

$$U = [D_{min} - D_1, D_{max} + D_2) = [0 - 0, 46.3287 + 0.6713) = [0, 47). \quad (26)$$

The union of sets U is divided into n intervals with the same interval length using the Sturges formula so that the value is obtained $n = 9$ and interval length of $l = 5.2222$. The obtained intervals are $u_1 = [0.0000, 5.2222)$, $u_2 = [5.2222, 10.4444)$, $u_3 = [10.4444, 15.6667)$, $u_4 = [15.6667, 20.8889)$, $u_5 = [20.8889, 26.1111)$, $u_6 = [26.1111, 31.3333)$, $u_7 = [31.3333, 36.5556)$, $u_8 = [36.5556, 41.7778)$, and

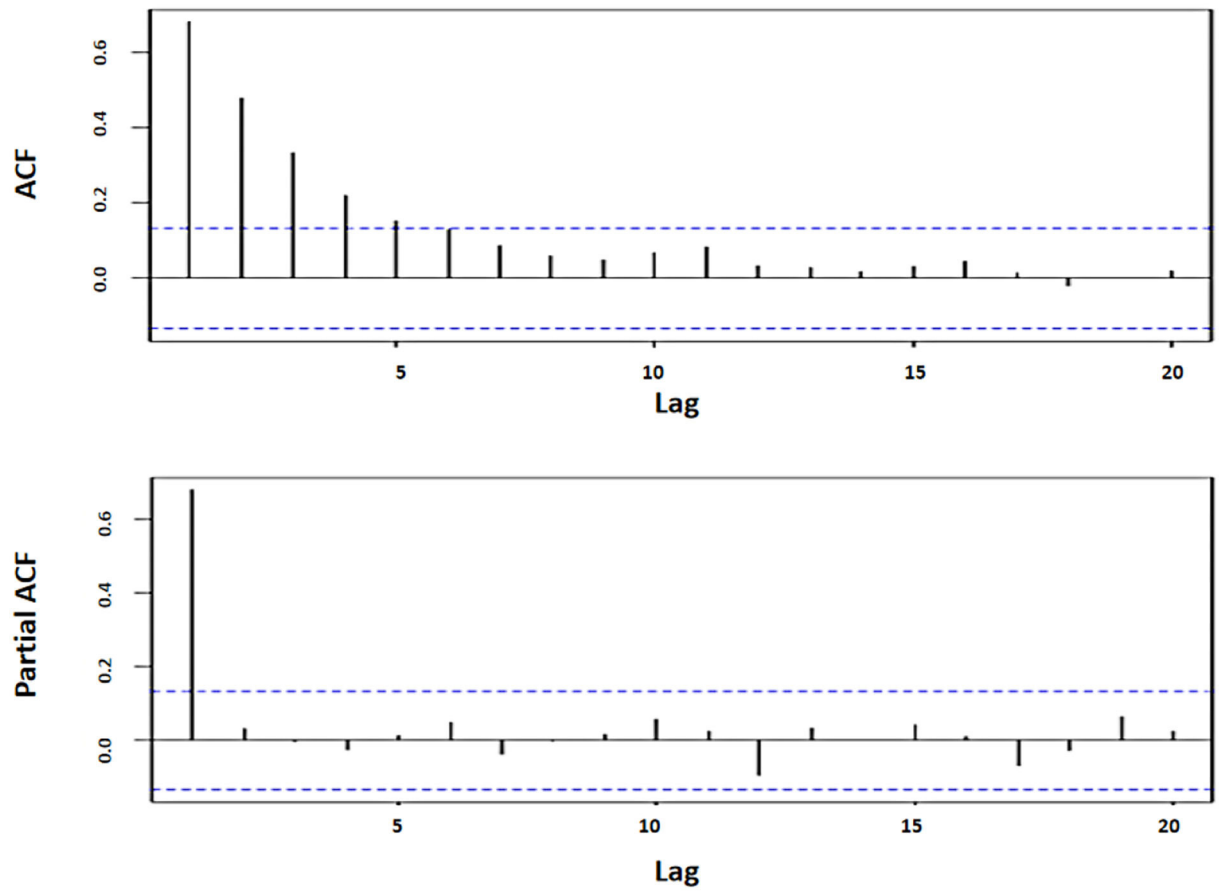


FIGURE 5
ACF and PACF plots of WTI oil data.

$u_9 = [41.7778, 47.0000)$. The next step is transforming the residual data into linguistic values, which are also in the form of intervals. Suppose the residual transformation data at the time $t = 1$ is 29.3201. It goes into the interval u_6. As a result, the transformation data at $t = 1$ after fuzzyfication are A_6.

Then, in the fuzzy logic relation (FLR) step, there is a relationship between the order of the data to the transformed data. This relationship is expressed by $A_i \rightarrow A_j$, where A_i is the left-hand side (LHS), and A_j is the right-hand side (RHS) from FLR. For example, FLR on a fuzzy set with value A_4 has a displacement relationship to A_4 or $A_4 \rightarrow A_4$. Therefore, the appearance of FLR $A_4 \rightarrow A_4$ is 3 on the data obtained so that fuzzy logic relation group (FLRG) from A_4 to A_4 can be written as $A_4 \rightarrow 3(A_4)$. FLRG for all data is shown in Table 8.

The fuzzy set relation above shows that FLR is in one group. The purpose of the relation states that the fuzzy set on the left side only has a relationship with the fuzzy set on the right side. For example, by using FLRG in Table 8, the transition probability matrix **R**, which has order 9, can be obtained as follows:

$$\mathbf{R} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ \frac{1}{12} & 0 & 0 & \frac{3}{12} & \frac{2}{12} & \frac{2}{12} & \frac{2}{12} & \frac{1}{12} & 0 \\ 0 & \frac{1}{42} & 0 & \frac{2}{42} & \frac{8}{42} & \frac{11}{42} & \frac{12}{42} & \frac{7}{42} & \frac{1}{42} \\ 0 & 0 & 0 & \frac{2}{69} & \frac{11}{69} & \frac{25}{69} & \frac{25}{69} & \frac{4}{69} & \frac{2}{69} \\ 0 & 0 & \frac{1}{65} & \frac{4}{65} & \frac{13}{65} & \frac{20}{65} & \frac{22}{65} & \frac{4}{65} & \frac{1}{65} \\ 0 & \frac{1}{20} & 0 & \frac{1}{20} & \frac{6}{20} & \frac{7}{20} & \frac{2}{20} & \frac{3}{20} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{3}{4} & \frac{1}{4} & 0 & 0 \end{bmatrix}. \quad (27)$$

The next step is to calculate the initial modeling value using the R matrix above. Using Equation 15, the initial modeling values for the residual transformation data at $t = 2$ are as

TABLE 5 Parameter estimation ARFIMA(p,d,q) model with $d = 0.4943$.

Model	ϕ	θ_1	θ_2	θ_3	θ_4	θ_5
ARFIMA(0,d,1)	-	−0.5642	-	-	-	-
		(< 2.2e−16)				
ARFIMA(0,d,2)	-	−0.6915	−0.3159	-	-	-
		(< 2.2e−16)	(1.1630e−8)			
ARFIMA(0,d,3)	-	−0.6858	−0.4208	-0.2067	-	-
		(< 2.2e−16)	(1.6827e−10)	(0.0023)		
ARFIMA(0,d,4)	-	-0.7203	-0.4740	-0.2970	0.1605	-
		(< 2.2e−16)	(8.3659e−9)	(9.5250e−5)	(0.0105)	
ARFIMA(0,d,5)	-	−0.7199	−0.4739	−0.2980	−0.1638	−0.0062
		(< 2.2e−16)	(9.3468e−9)	(0.0001)	(0.0297)	(0.9369)
ARFIMA(1,d,0)	0.6928	-	-	-	-	-
	(< 2.2e−16)					
ARFIMA(1,d,1)	0.6779	−0.0292	-	-	-	-
	(< 2e−16)	(0.7681)				
ARFIMA(1,d,2)	0.6883	−0.0185	0.0134			
	(2.3446e−11)	(0.8819)	(0.8923)			
ARFIMA(1,d,3)	0.6968	-0.0101	0.0201	0.0067		
	(9.6578e−6)	(0.9528)	(0.8836)	(0.9465)		
ARFIMA(1,d,4)	0.6748	-0.0331	0.0052	−0.0036	-0.0120	
	(0.0128)	(0.9081)	(0.9795)	(0.9801)	(0.9179)	
ARFIMA(1,d,5)	0.8916	0.1787	0.1609	0.1040	0.06590	0.0936
	(3.0606×10^{-7})	(0.3406)	(0.2738)	(0.3668)	(0.5193)	(0.2803)

follows:

$$\begin{aligned} F_2 &= m_4P_{64} + m_5P_{65} + X_1P_{66} + m_7P_{67} + m_8P_{68} + m_9P_{69} \\ &= (18.2778)\left(\frac{2}{69}\right) + (23.5000)\left(\frac{11}{69}\right) + (29.3201)\left(\frac{25}{69}\right) \\ &+ (33.9444)\left(\frac{25}{69}\right) + (39.1667)\left(\frac{4}{69}\right) + (44.3889)\left(\frac{2}{69}\right) \\ &= 30.7553. \end{aligned}$$

Then the modeling adjustment value for the residual transformation data at $t = 2$ is

$$D_t = \left(\frac{l}{2}\right)s = \left(\frac{5.2222}{2}\right)(1) = 2.6111.$$

Final modeling values for residual transformation data at time $t = 2$ are:

$$FM_2 = F_2 + D_t = 30.7553 + 2.6111 = 33.3664.$$

Transformation returns to return the final modeling residual transformation data to the initial modeling residual data. This is done by adding the minimum residual data to the final modeling residual transformation data. For example when $t = 2$, the final modeling value of residual data with FTSMC is $33.3664 + (-29.2866) = 4.0798$. The results of the ARIMA-FTSMC hybrid modeling were obtained by adding the ARIMA(1,1,0) model data with the final ARIMA(1,1,0) residual modeling data. For example, when $t = 2$, the hybrid modeling value is $33.5996 + 4.0798 = 37.6794$.

TABLE 6 Comparison of AIC and BIC values ARFIMA (p,d,q) model with $d = 0.4943$.

Model	AIC	BIC
ARFIMA(0,d,1)	−3.388.140	−3.252.950
ARFIMA(0,d,2)	−3.657.100	−3.488.110
ARFIMA(0,d,3)	−3.728.870	−3.526.080
ARFIMA(0,d,4)	−3.771.410	−3.534.810
ARFIMA(1,d,0)	−3.842.370	−3.707.170

TABLE 7 Residual assumption test ARFIMA(1,d,0) model with $d = 0.4943$.

Residual assumption	Non-autocorrelation	Normality
p-value	0.8271	0.0627

TABLE 8 The fuzzy logic relation group (FLRG).

No	FLRG
1	$A_1 \rightarrow A_5$
2	$A_2 \rightarrow A_5, A_7$
3	$A_3 \rightarrow A_7$
4	$A_4 \rightarrow A_1, 3(A_4), 2(A_5), 2(A_6), 2(A_7), 2(A_8)$
5	$A_5 \rightarrow A_2, 2(A_4), 8(A_5), 11(A_6), 12(A_7), 7(A_8), A_9$
6	$A_6 \rightarrow 2(A_4), 11(A_5), 25(A_6), 25(A_7), 4(A_8), 2(A_9)$
7	$A_7 \rightarrow A_3, 4(A_4), 13(A_5), 20(A_6), 22(A_7), 4(A_8), A_9$
8	$A_8 \rightarrow A_2, A_4, 6(A_5), 7(A_6), 2(A_7), 3(A_8)$
9	$A_9 \rightarrow 3(A_6), A_7$

TABLE 9 The fuzzy logic relation group (FLRG).

No	FLRG
1	$A_1 \rightarrow A_5, A_6$
2	$A_2 \rightarrow A_7$
3	$A_3 \rightarrow A_4, A_5, A_6, A_7$
4	$A_4 \rightarrow A_1, 2(A_3), 3(A_5), 4(A_7), A_8, A_9$
5	$A_5 \rightarrow A_1, A_3, 3(A_4), 12(A_5)), 17(A_6), 14(A_7), 7(A_8))$
6	$A_6 \rightarrow 3(A_4), 17(A_5)), 31(A_6)), 16(A_7)), A_8, 3(A_9)$
7	$A_7 \rightarrow A_2, 5(A_4), 14(A_5), 15(A_6), 12(A_7), 4(A_8)$
8	$A_8 \rightarrow 6(A_5), 5(A_6), 2(A_7), A_8$
9	$A_9 \rightarrow A_3, A_5, A_6, A_7, 2(A_9)$

3.4. The hybrid ARFIMA-FTSMC model of WTI oil price

The hybrid ARFIMA-FTSMC model is a combination of the ARFIMA and FTSMC models. The first step is to build a model of time series data with the ARFIMA model. The results show that the ARFIMA(1,d,0) model is selected to be the best model. Thus, the residual data of the ARFIMA(1,d,0) model, denoted by ϵ_t, is built using the FTSMC model. In the form of a negative number, the transformation is carried out so that the data becomes a non-negative number. The data transformation is used to determine the minimum data from the residual data. Then, residual data are subtracted from the minimum data to become a non-negative number. Based on the residual data, the result shows the minimum residual data is $d_{min} = -26.8005$. suppose the residual data at the time t=3, which is $\epsilon_3 = -6.2003$ so that the residual transformation data obtained is 20.6002. The first step in modeling residual data using the FTSMC model is to determine the union of set U. Based on the residual transformation data. It is obtained that $D_{max} = 43.6958$, $D_{min} = 0$ and the determination of value $D_1 = 0$, $D_2 = 1.3042$. The union of set U is defined as an interval below:

$$U = [D_{min} - D_1; D_{max} + D_2) = [0 - 0; 43.6958 + 1.3042) = [0; 45)$$

The union of sets U is divided into n intervals with the same interval length using the Sturges formula so that the value is obtained n=9 and interval length of $l = 5$. The obtained intervals are $u_1 = [0, 5)$, $u_2 = [5, 10)$, $u_3 = [10, 15)$, $u_4 = [15, 20)$, $u_5 = [20, 25)$, $u_6 = [25, 30)$, $u_7 = [30, 35)$, $u_8 = [35, 40)$, $u_9 = [40, 45)$. The next step is transforming the residual data into linguistic values, which are also in the form of intervals. Suppose the residual transformation data at the time $t = 2$ is 29.7330 it goes to interval u_6. As a result, the transformation data at time $t = 2$ after fuzzyfication is A_6.

Then, in the fuzzy logic relation (FLR) step, there is a relationship between the order of the data to the transformed data. This relationship is expressed by $A_i \rightarrow A_j$, where A_i is the left-hand side (LHS), and A_j is the right-hand side (RHS) from FLR. For example, FLR on a fuzzy set with value A_4 has a displacement relationship to A_4 or $A_4 \rightarrow A_4$. Therefore, the appearance of FLR $A_4 \rightarrow A_4$ is 3 on the data obtained so that the fuzzy logic relation group (FLRG) from A_4 to A_4 can be written as $A_4 \rightarrow 3(A_4)$. FLRG for all data is shown in Table 9:

The fuzzy set relation above shows that FLR is in one group. The purpose of the relation states that the fuzzy set on the left side only has a relationship with the fuzzy set on the right side. For example, by using FLRG in Table 8, the transition probability matrix **R**, which has order 9, can be obtained as follows:

$$\mathbf{R} = \begin{bmatrix} 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 0 & 0 \\ \frac{1}{12} & 0 & \frac{2}{12} & 0 & \frac{3}{12} & 0 & \frac{4}{12} & \frac{1}{12} & \frac{1}{12} \\ \frac{1}{55} & 0 & \frac{1}{55} & \frac{3}{55} & \frac{12}{55} & \frac{17}{55} & \frac{14}{55} & \frac{7}{55} & 0 \\ 0 & 0 & 0 & \frac{3}{71} & \frac{17}{71} & \frac{31}{71} & \frac{16}{71} & \frac{1}{71} & \frac{3}{71} \\ 0 & \frac{1}{51} & 0 & \frac{5}{51} & \frac{14}{51} & \frac{15}{51} & \frac{12}{51} & \frac{4}{51} & 0 \\ 0 & 0 & 0 & 0 & \frac{6}{14} & \frac{5}{14} & \frac{2}{14} & \frac{1}{14} & 0 \\ 0 & 0 & \frac{1}{6} & 0 & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & 0 & \frac{2}{6} \end{bmatrix}. \quad (28)$$

The next step is to calculate the initial modeling value using the R matrix above. The initial modeling values for the residual

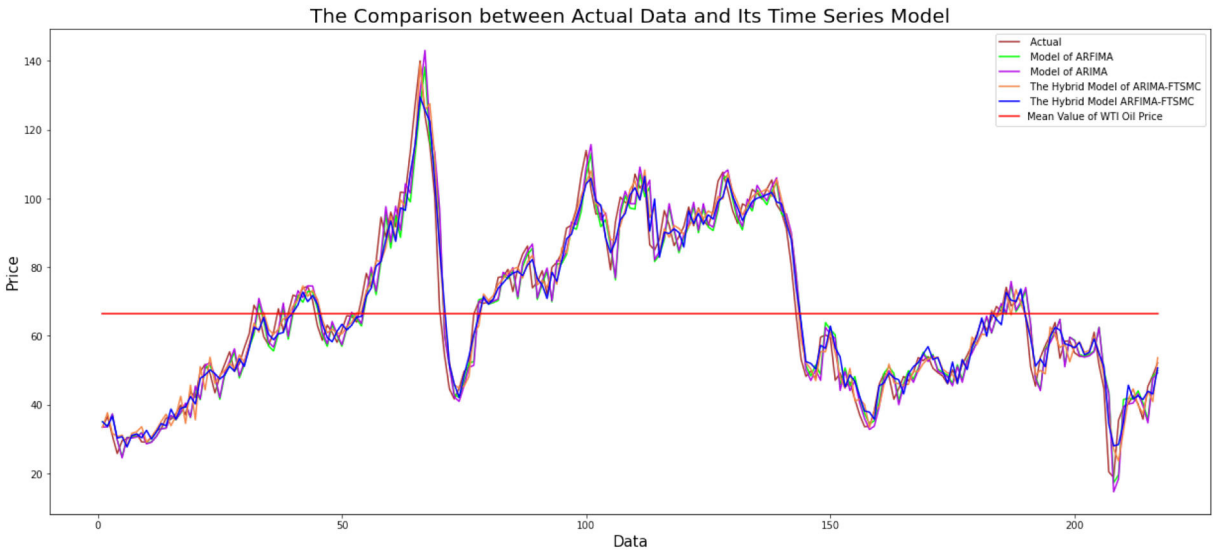


FIGURE 6
Comparison graph of actual data with ARIMA, ARFIMA, hybrid ARIMA-FTSMC, and hybrid ARFIMA-FTSMC models.

transformation data at $t = 4$ are as follows:

$$\begin{aligned} F_4 &= m_1P_{51} + m_3P_{53} + m_4P_{54} + X_3P_{55} + m_6P_{56} + m_7P_{57} + m_8P_{58} \\ &= (2.5)\frac{1}{55} + (12.5)\frac{1}{55} + (17.5)\frac{3}{55} + (20.6002)\frac{12}{55} + (27.5)\frac{17}{55} \\ &\quad + (32.5)\frac{14}{55} + (37.5)\frac{7}{55} \\ &= 27.2673. \end{aligned}$$

Then the modeling adjustment value for the residual transformation data at $t = 2$ is

$$D_t = \left(\frac{l}{2}\right)s = \left(\frac{5}{2}\right)(0) = 0.$$

Final modeling values for residual transformation data at time $t = 2$ are:

$$FM_4 = F_4 + D_t = 27.2673 + 0 = 27.2673.$$

Transformation return to return the final modeling residual transformation data to the initial modeling residual data. This is done by adding the minimum residual data to the final modeling residual transformation data. For example when t=4, the final modeling value of residual data with FTSMC is 27.2673+(-26.8005)=0.4668. The results of the ARFIMA-FTSMC hybrid modeling were obtained by adding the ARFIMA(1,d,0) model data with the final ARFIMA(1,d,0) residual data. For example, when t=4, the hybrid modeling value is 29.8487+0.4668=30.3156.

3.5. The comparison and discussion

After building the model of ARIMA, ARFIMA, a hybrid model of ARIMA-FTSMC, and a hybrid model of ARFIMA-FTSMC, then, we continue to do comparisons among the models by analyzing graphical and accuracy measurements. The following is a graphical comparison of the ARIMA, ARFIMA, hybrid ARIMA-FTSMC, and hybrid ARFIMA-FTSMC models based on the WTI data presented in Figure 6 below:

From Figure 6, it can be seen that modeling using ARIMA, ARFIMA, hybrid ARIMA-FTSMC, and hybrid ARFIMA-FTSMC models provide modeling results that are close to the actual data. Although graphically, the three models show good estimation results, it is necessary to examine the level of accuracy of each model to see a more precise model with better accuracy that can be used in predictions.

Calculation of modeling accuracy to see the preferred model for modeling WTI oil price data from the three models is presented using MAE, RMSE, and MAPE values. By using Equations (22)–(24), the modeling accuracy is shown in Table 10.

Table 10 shows the accuracy measures, the best selection criteria, and the goodness of fit for each model. Based on the MAE and RMSE values, all models have accuracy values that are not too far off. When viewed from the MAPE criteria, all models have a MAPE value of less than 10%. This means that all models can model WTI oil data very well. Based on the accuracy of the model, the ARFIMA-FTSMC has the best performance than the rest model, and so does the best selection and the

TABLE 10 The comparison of the ARIMA model with GARCH, FFNN, and GARCH-FFNN using model accuracy, the best model selection, and goodness of fit R^2.

Model	Model accuracy			The best model selection measurement			R^2
	MAE	RMSE	MAPE	Log-Likelihood	AIC	BIC	
ARIMA	4.8201	6.3901	7.9257	3.7095	806.9615	810.3414	0.9285
ARIMA-FTSMC	3.6533	4.7863	6.1219	3.1315	681.5355	684.9154	0.9608
ARFIMA	4.7973	6.3275	7.8612	3.6898	802.6866	806.0665	0.9299
ARFIMA-FTSMC	3.5268	4.5777	6.0762	3.0424	662.2008	665.5807	0.9633

goodness of fit criterion. Table 10 shows that the ARFIMA-FTSMC has the smallest value of MAE, RMSE, MAPE, Log-likelihood, AIC, and BIC. Then, the second one is ARIMA-FTSMC which has the model accuracy and goodness of fit that is closer to ARFIMA-FTSMC. It shows that if the FTSMC model adjusts the residual model of ARIMA or ARFIMA then the accuracy and goodness of fit of the new proposed model will be better than the model without improvements using FTSMC. Therefore, the hybrid model of ARFIMA-FTSMC gives better results in modeling monthly WTI oil prices than the rest models.

4. Conclusion

Based on the data analysis that has been done, it can be seen that the WTI oil data have a long-memory data pattern. This value is detected from the ACF plot, which decreases slowly over time. This shows the influence of past data on the current data is still strong and will decrease over time. Therefore, time series data for WTI oil prices can be formed into the model of ARIMA, the model of ARFIMA, the hybrid model of ARIMA-FTSMC, and the hybrid model of ARFIMA-FTSMC. The four models provide reasonable parameter estimates and high accuracy, which shows their closeness to the actual data. This is obtained based on the accuracy value of each model obtained not too far from the actual data. However, the proposed new model, the hybrid model of ARFIMA-FTSMC, provides the smallest MAE, RMSE, and MAPE values. It can be concluded that the hybrid model of ARFIMA-FTSMC has better accuracy than other models.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: https://www.eia.gov/.

Author contributions

DD contributed to the conception of the hybrid model and design of the study. KR played a role in analyzing the model of the fuzzy time series (FTS) and wrote sections of the article. M and YA performed the time series analysis of ARIMA and ARFIMA, respectively. MY organized the database and numerical simulation of data processing. All authors contributed to the article revision, read, and approved the submitted version.

Acknowledgments

The authors acknowledge Andalas University under the Ministry of Education, Cultural, Research and Technology for funding this research under the scheme of university basic research excellency with contract number 011/E5/PG.02.00.PT.2022.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Wei WWS. *Multivariate Time Series Analysis and Applications*. New York, NY: Wiley (2019).

2. Ramadani K, Devianto D. The forecasting model of Bitcoin price with fuzzy time series Markov chain and chen logical method. *AIP Conf Proc*. (2020) 2296:1–11. doi: 10.1063/5.0032178

3. Devianto D, Maiyastri, Fadhilla DR. Time series modeling for risk of stock price with value at risk computation. *Appl Math Sci*. (2015) 9:2779–87. doi: 10.12988/ams.2015.52144

4. Mohammad AA, Mudhir AA. Dynamical approach in studying stability condition of exponential (GARCH) models. *J King Saud Univer Sci*. (2020) 32:272–8. doi: 10.1016/j.jksus.2018.04.028

5. Zeghdoudi H, Amrani M. On mixture GARCH models: long, short memory and application in finance. *J Math Stat Stud*. (2021) 2:01–07. doi: 10.32996/jmss.2021.2.2.1

6. Park SH, Griffiths JD, Longtin A, Lefebvre J. Persistent entrainment in non-linear neural networks with memory. *Front Appl Math Stat*. (2018) 4:31. doi: 10.3389/fams.2018.00031

7. Garnier J, K Solna K. Implied volatility structure in turbulent and long-memory markets. *Front Appl Math Stat*. (2020) 6:10. doi: 10.3389/fams.2020.00010

8. Baillie RT, Morana C. Adaptive ARFIMA models with applications to inflation. *Econ Model*. (2012) 29:2451–9. doi: 10.1016/j.econmod.2012.07.011

9. Baillie RT, Kongcharoen C, Kapetanios G. Prediction from ARFIMA models: comparisons between MLE and semiparametric estimation procedures. *Int J Forecast*. (2012) 28:46–53. doi: 10.1016/j.ijforecast.2011.02.012

10. Zhang X, Lu Z, Wang Y, Zhang R. Adjusted jackknife empirical likelihood for stationary ARMA and ARFIMA models. *Stat Probab Lett*. (2020) 165:1–11. doi: 10.1016/j.spl.2020.108830

11. Boaisha SM, Amaitik SM. Forecasting model based on fuzzy time series approach. In: *Proceedings of the 10th International Arab Conference on Information Technology-ACIT*. (2010). p. 14–16.

12. Efendi R, Deris MM. Forecasting of malaysian oil production and oil consumption using fuzzy time series. *Recent Adv Soft Comput Data Min*. (2016) 441:31–40. doi: 10.1007/978-3-319-51281-5_4

13. Tsaur RC. A fuzzy time series-markov chain model with an application to model the exchange rate between the taiwan and US dollar. *Int J Innovat Comput Inf Control*. (2012) 8:4931–42.

14. Uzun B, KÄśral E. Application of Markov chains-fuzzy states to gold price. *Procedia Comput Sci*. (2017) 120:365–71. doi: 10.1016/j.procs.2017.11.251

15. Gao R, Duru O. Parsimonious fuzzy time series modelling. *Expert Syst Appl*. (2020) 156:1–12. doi: 10.1016/j.eswa.2020.113447

16. Yollanda M, Devianto D, Yozza H. Nonlinear modeling of IHSG with artificial intelligence. In: *2018 International Conference on Applied Information Technology and Innovation*. Padang: IEEE Xplore (2018). p. 85–90.

17. Adib A, Zaerpour A, Lotfirad M. On the reliability of a novel MODWT-based hybrid ARIMA-artificial intelligence approach to forecast daily snow depth (Case study: the western part of the Rocky Mountains in the USA). *Cold Regions Sci Technol*. (2021) 189:1–11. doi: 10.1016/j.coldregions.2021.103342

18. Denis DJ. *Applied Univariate, Bivariate, and Multivariate Statistics: Understanding Statistics for Social and Natural Scientists, with Applications in SPSS and R*. New York, NY: Wiley (2021).

frontiers | Frontiers in Applied Mathematics and Statistics

TYPE Original Research
PUBLISHED 28 February 2023
DOI 10.3389/fams.2023.1029018

OPEN ACCESS

EDITED BY
Joel M. Addawe,
University of the Philippines Baguio, Philippines

REVIEWED BY
Appanah Rao Appadu,
Nelson Mandela University, South Africa
Afifurrahman -,
Universitas Islam Negeri Mataram, Indonesia

*CORRESPONDENCE
Randy L. Caga-anan
✉ randy.caga-anan@g.msuiit.edu.ph

SPECIALTY SECTION
This article was submitted to Mathematics of Computation and Data Science, a section of the journal Frontiers in Applied Mathematics and Statistics

RECEIVED 26 August 2022
ACCEPTED 31 January 2023
PUBLISHED 28 February 2023

CITATION
Campos HJ, Raza MN, Arcede JP, Martinez JGT and Caga-anan RL (2023) Vaccination and variants: A COVID-19 multi-strain model evolution for the Philippines.
Front. Appl. Math. Stat. 9:1029018.
doi: 10.3389/fams.2023.1029018

Vaccination and variants: A COVID-19 multi-strain model evolution for the Philippines

Harren J. Campos[1], Michelle N. Raza[2], Jayrold P. Arcede[3], Joey Genevieve T. Martinez[4,5] and Randy L. Caga-anan[5,6]*

[1]Mathematics Department, Mindanao State University-Main Campus, Marawi City, Philippines, [2]Division of Natural Sciences and Mathematics, University of the Philippines Visayas Tacloban College, Tacloban City, Philippines, [3]Mathematics Department, Caraga State University, Butuan City, Philippines, [4]Department of Biological Sciences, Mindanao State University-Iligan Institute of Technology, Iligan City, Philippines, [5]Mathematical Biology and Nematology Research Cluster, Complex Systems Group, Premier Research Institute of Science and Mathematics, Mindanao State University-Iligan Institute of Technology, Iligan City, Philippines, [6]Department of Mathematics and Statistics, Mindanao State University-Iligan Institute of Technology, Iligan City, Philippines

Coronavirus disease 2019 (COVID-19) management and response is a challenging task due to the uncertainty and complexity of the nature surrounding the virus. In particular, the emergence of new variants and the polarizing response from the populace complicate government efforts to control the pandemic. In this study, we developed a compartmental model that includes (1) a vaccinated compartment, (2) reinfection after a particular time, and (3) COVID-19 variants dominant in the Philippines. Furthermore, we incorporated stochastic terms to capture uncertainty brought about by the further evolution of the new variants and changing control measures *via* parametric perturbation. Results show the importance of booster shots that increase the vaccine-induced immunity duration. Without booster shots, simulations showed that the dominant strain would still cause significant infection until 31 December 2023. Moreover, our stochastic model output showed significant variability in this case, implying greater uncertainty with future predictions. All these adverse effects, fortunately, can be effectively countered by increasing the vaccine-induced immunity duration that can be done through booster shots.

KEYWORDS

COVID-19 variants, Philippines, vaccination, mathematical model, stochastic simulation, multi-strain

1. Introduction

Since the emergence of coronavirus disease 2019 (COVID-19) in Wuhan, its causative agent, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has been rapidly evolving into new variants [1]. This emergence is expected as, like any virus, SARS-CoV-2 continues to mutate from time to time. So far, the World Health Organization (WHO) has identified five variants of concern, namely, Alpha, Beta, Gamma, Delta, and Omicron. These are the variants considered the most transmissible and dominant that are circulating the world [2]. While vaccine has been available, the campaign has suffered a series of setbacks and there are many issues leading to vaccine hesitancy [3, 4]. In the Philippines, the slow vaccine rollout, limited testing capacity, weak genomics surveillance, a fragile healthcare system, and a large informal economy contributed more to the many issues related to COVID-19. The latter is a social cost as few people can afford not to work but are forced to stay home in overcrowded housing. These factors exacerbated the situation as they gave the virus a perfect environment to mutate, feeding through a continuous supply of susceptibles.

Hence, a question of much importance is how to reduce COVID-19 transmission. In doing so, an approach of balancing the efforts on the vaccination campaign (booster shots amidst reinfection because of mutating variants and waning immunity), and preparedness of the healthcare system coupled with public health and social measures is a must. While the government is trying to strike a balance, the race between vaccination and fast mutating variants remains the biggest challenge.

Many computational modeling studies have investigated COVID-19 dynamics in the Philippines. For instance, in Arcede et al. [5], an SEIR-type model is constructed whose infected can either be symptomatic or not. The result shows that treating symptomatic alone does not reduce the spread. However, managing the number of susceptible does, which containment and vaccines have a significant impact role to play. Later, the same model was used to investigate the implemented non-pharmaceutical interventions (NPI) in the country [6]. Here, NPIs include lockdown, social distancing, mass testing, and strengthening the healthcare system. The study provided a choice for the government to implement the control by indicating economic cost (low, high) given no vaccine availability. In Bock et al. [7], an agent-based SIR model was used to investigate the prevalence of COVID-19 in two neighboring cities in Northern Mindanao, particularly in Iligan and Cagayan de Oro. The result shows that social distancing and age-specific quarantine can effectively slow down contagion. Furthermore, social distancing combined with an effective testing strategy can keep the epidemic at bay and prevent it from becoming a critical epidemic. In Arcede et al. [8], a regional COVID-19 model has been constructed in the cities mentioned earlier and in the Northern Mindanao region as a whole. The model is tailored to fit early transmission; hence appropriate models are suggested when laboratory-based disease reports are available. In [9], Mammeri et al. extended their SEIR-type model to account the spatial movement of individuals. Given five main islands and five main airports in the Philippines as nodes with index case assumed to start from Manila airport on day 1, their simulation show remarkably close similarity to what happened in the Philippines during its first 140 days. Studies mentioned do not deal with vaccination control. However, recent articles deal with vaccination strategies in the Philippine context. For instance, in [10], optimal control was used to investigate existing policy interventions, including vaccination rollouts, community quarantines, and simulated virus outbreaks. They found that early and effective implementation of precautionary measures such as community quarantines are crucial for containing outbreaks. They also found that even if vaccinations do not suffice, expanding the vaccine supply reduces the need for more resource-intensive interventions. Moreover, in Caga-anan et al. [11], a model with a delay on the vaccination compartment was constructed to study the impact of vaccination efforts on disease progression and herd immunity. The result shows that timely vaccination is preferred to maximize impact. They also assessed the performance of different vaccine brands in the model, showing Pfizer-BioNTech with the best results. Finally, some models were proposed for allocating resources. For instance, optimizing the location of vaccination sites implemented in San Juan Philippines [12] and distribution of COVID-19 testing kits in DOH-accredited testing centers in the country [13]. All studies mentioned earlier do not account explicitly for COVID-19 variants and randomness.

In this study, we evaluate the impact of vaccination and its waning induced immunity, as well as the uncertainties related to future mitigation policies and the further evolution of the virus, given the three variants that are circulating dominantly in the country.

2. Model formulation

2.1. Deterministic model

In this study, we divided the population into eight compartments, namely, susceptible (S), vaccinated (V), infected by the original strain (I_1), infected by the Delta variant (I_2), infected by the Omicron variant (I_3), confirmed (C), recovered (R), and dead (D). Figure 1 shows the dynamics of the model. The model considers the following facts. First, vaccines do not provide lasting immunity. Hence, vaccinated people will be back to being susceptible after some time. Second, reported confirmed cases do not show the full extent of the infection. Third, estimates of unreported cases are not being accounted for. Finally, recovered individuals (from natural infection) also do not have lasting immunity against the virus; hence, both vaccinated and recovered individuals may return to the susceptible population after some time. In the model, we assume density-dependent transmission rates.

An infected individual by any of the variants can infect people in compartment S. We assume that vaccinated people in V are immune to any of the variants. The parameter β_i denotes the transmission rate of the disease caused by the ith strain. The parameter α represents the level of control measures implemented to limit the transmission. The parameter ν denotes the vaccination rate of susceptible people. The parameters λ_1 and λ_2 denote the vaccine immunity waning rate and natural immunity waning rate, respectively. The parameter ϵ denotes the proportion of infections detected and confirmed through testing also known as detection rate. We also have the removal rate from C given by δ. Finally, the recovery rate from unconfirmed infections caused by the ith strain is denoted as γ_i and the probability to recover from the infection is denoted as ρ. The parameters are summarized in Table 1.

In this study, we adapted a closed population model, i.e., we did not consider natural birth and death rates. We denote by N_0 the total number of population at the start. The dynamics of our model is governed by the following system of ordinary differential equations (ODE):

$$\frac{dS}{dt} = -\alpha(\beta_1 I_1 + \beta_2 I_2 + \beta_3 I_3)\frac{S}{N_0} - \nu S + \lambda_1 V + \lambda_2 R \quad (1)$$

$$\frac{dV}{dt} = \nu S - \lambda_1 V \quad (2)$$

$$\frac{dI_1}{dt} = \alpha\beta_1 I_1 \frac{S}{N_0} - (\gamma_1 + \epsilon) I_1 \quad (3)$$

$$\frac{dI_2}{dt} = \alpha\beta_2 I_2 \frac{S}{N_0} - (\gamma_2 + \epsilon) I_2 \quad (4)$$

$$\frac{dI_3}{dt} = \alpha\beta_3 I_3 \frac{S}{N_0} - (\gamma_3 + \epsilon) I_3 \quad (5)$$

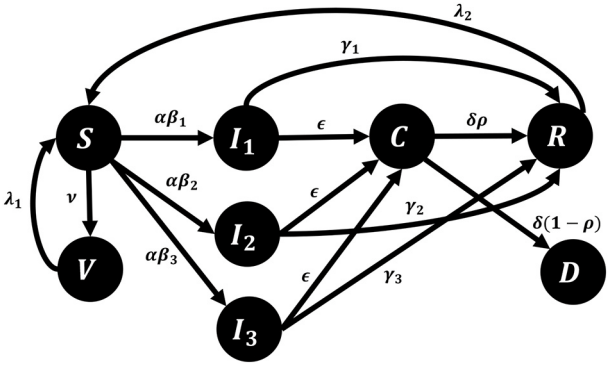


FIGURE 1
Model dynamics of COVID-19 with three strains.

TABLE 1 Parameters of the model.

Parameter	Description	Unit
β_i	Transmission rate due to I_i	1/day
α	Transmission reduction	Dimensionless
ν	Vaccination rate	1/day
λ_1	Vaccine immunity waning rate	1/day
λ_2	Natural immunity waning rate	1/day
ϵ	Detection rate	1/day
γ_i	Recovery rate from I_i	1/day
δ	Removal rate from C	1/day
ρ	Probability to recover	Dimensionless

$$\frac{dC}{dt} = \epsilon I_1 + \epsilon I_2 + \epsilon I_3 - \delta C \tag{6}$$
$$\frac{dR}{dt} = \gamma_1 I_1 + \gamma_2 I_2 + \gamma_3 I_3 + \delta\rho C - \lambda_2 R \tag{7}$$
$$\frac{dD}{dt} = \delta(1-\rho)C. \tag{8}$$

2.2. Stochastic model

As will be seen in Section 5, the third strain will become dominant as time goes on. We acknowledge that we need to accommodate future uncertainties in our simulation. Hence, we modify our ODE model to add stochasticity. We apply parametric perturbation to the reduced transmission rate $\alpha\beta_3$. The resulting system with stochastic differential equations (SDE) is as follows:

$$\begin{aligned}
dS &= \left[-\alpha(\beta_1 I_1 + \beta_2 I_2 + \beta_3 I_3)\frac{S}{N_0} - \nu S + \lambda_1 V + \lambda_2 R\right] dt \\
&\quad -\sigma I_3 \frac{S}{N_0} dB \\
dV &= [\nu S - \lambda_1 V]\, dt \\
dI_1 &= \left[\alpha\beta_1 I_1 \frac{S}{N_0} - (\gamma_1 + \epsilon) I_1\right] dt \\
dI_2 &= \left[\alpha\beta_2 I_2 \frac{S}{N_0} - (\gamma_2 + \epsilon) I_2\right] dt \qquad (9) \\
dI_3 &= \left[\alpha\beta_3 I_3 \frac{S}{N_0} - (\gamma_3 + \epsilon) I_3\right] dt + \sigma I_3 \frac{S}{N_0} dB \\
dC &= [\epsilon I_1 + \epsilon I_2 + \epsilon I_3 - \delta C]\, dt \\
dR &= [\gamma_1 I_1 + \gamma_2 I_2 + \gamma_3 I_3 + \delta\rho C - \lambda_2 R]\, dt \\
dD &= [\delta(1-\rho)C]\, dt,
\end{aligned}$$

where dB/dt is the white noise, i.e., the derivative of the standard Brownian motion $B(t)$, and $\sigma > 0$ denotes the intensity of that noise.

3. Qualitative analysis of the ODE model

Since we have a close system and $N_0 = S + V + I_1 + I_2 + I_3 + C + R + D$, we may just consider the system (1)–(7), i.e., without D. To find the disease-free equilibrium (DFE) (a steady-state solution of an epidemic model with all infected variables equals to zero), we equate Equations (1)–(7) to zero and the infective compartments I_1, I_2, I_3, and C equal to 0. We obtain a solution, which is the DFE given by $(S^*, V^*, 0, 0, 0, 0, 0)$, where $S^* = \frac{\lambda_1 V^*}{\nu}$.

3.1. Reproduction number

By definition, the basic reproduction number $\mathcal{R}_0$ denotes the average number of individuals directly infected by a single infected individual over the duration of its infectious period in a population without any deliberate intervention to stop its spread. We will compute $\mathcal{R}_0$ using the next generation matrix method defined by Diekman et al. [14] and van den Driessche and Watmough [15]. Let X be the vector of the infected classes and Y be the vector of the other classes. Let $\mathcal{F}(X, Y)$ be the vector of new infection rates (flows from Y to X) and let $\mathcal{V}(X, Y)$ be the vector of all other rates (not new infections). Then, we have

$$\mathcal{F} = \begin{bmatrix} \alpha\beta_1 I_1 \frac{S}{N_0} \\ \alpha\beta_2 I_2 \frac{S}{N_0} \\ \alpha\beta_3 I_3 \frac{S}{N_0} \\ \epsilon I_1 + \epsilon I_2 + \epsilon I_3 \end{bmatrix}, \mathcal{V} = \begin{bmatrix} (\gamma_1 + \epsilon) I_1 \\ (\gamma_2 + \epsilon) I_2 \\ (\gamma_3 + \epsilon) I_3 \\ \delta C \end{bmatrix}.$$

Evaluating the derivatives of $\mathcal{F}$ and $\mathcal{V}$ at the DFE, we are led to the following matrices

$$F = \begin{bmatrix} \alpha\beta_1 \frac{S^*}{N_0} & 0 & 0 & 0 \\ 0 & \alpha\beta_2 \frac{S^*}{N_0} & 0 & 0 \\ 0 & 0 & \alpha\beta_3 \frac{S^*}{N_0} & 0 \\ \epsilon & \epsilon & \epsilon & 0 \end{bmatrix}, V = \begin{bmatrix} \gamma_1 + \epsilon & 0 & 0 & 0 \\ 0 & \gamma_2 + \epsilon & 0 & 0 \\ 0 & 0 & \gamma_3 + \epsilon & 0 \\ 0 & 0 & 0 & \delta \end{bmatrix}.$$

Hence, the next generation matrix is given by

$$K = FV^{-1} = \begin{bmatrix} \left(\frac{\alpha\beta_1}{\gamma_1+\epsilon}\right)\frac{S^*}{N_0} & 0 & 0 & 0 \\ 0 & \left(\frac{\alpha\beta_2}{\gamma_2+\epsilon}\right)\frac{S^*}{N_0} & 0 & 0 \\ 0 & 0 & \left(\frac{\alpha\beta_3}{\gamma_3+\epsilon}\right)\frac{S^*}{N_0} & 0 \\ \frac{\epsilon}{\gamma_1+\epsilon} & \frac{\epsilon}{\gamma_2+\epsilon} & \frac{\epsilon}{\gamma_3+\epsilon} & 0 \end{bmatrix}.$$

The eigenvalues of K are the following: $\eta_1 = \left(\frac{\alpha\beta_1}{\gamma_1+\epsilon}\right)\frac{S^*}{N_0}$, $\eta_2 = \left(\frac{\alpha\beta_2}{\gamma_2+\epsilon}\right)\frac{S^*}{N_0}$, $\eta_3 = \left(\frac{\alpha\beta_3}{\gamma_3+\epsilon}\right)\frac{S^*}{N_0}$, and $\eta_4 = 0$. The eigenvalue η_1 is associated with strain 1 and gives rise to the basic reproduction number $\mathcal{R}_1 = \left(\frac{\lambda_1\alpha\beta_1}{\nu(\gamma_1+\epsilon)}\right)\frac{V^*}{N_0}$. Similarly, eigenvalue η_2 associated with strain 2 corresponds to the basic reproduction number $\mathcal{R}_2 = \left(\frac{\lambda_1\alpha\beta_2}{\nu(\gamma_2+\epsilon)}\right)\frac{V^*}{N_0}$, and eigenvalue η_3 associated with strain 3 corresponds to $\mathcal{R}_3 = \left(\frac{\lambda_1\alpha\beta_3}{\nu(\gamma_3+\epsilon)}\right)\frac{V^*}{N_0}$. Finally, we take $\mathcal{R}_0 = \max\{\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3\}$.

3.2. Stability analysis of the DFE

Theorem 3.1. *The disease-free equilibrium* $x_0 = (S^*, V^*, 0, 0, 0, 0, 0, D^*)$, *where* $S^* = \frac{\lambda_1 V^*}{\nu}$ *and* $D^* = N_0 - (S^* + V^*)$ *of system (1)-(8) is globally asymptotically stable.*

Proof. Consider that

$$D(t) - D(0) = \int_0^t \delta(1-\rho)\, C(s)\, ds \tag{10}$$

and

$$D^* - D(0) = \int_0^{+\infty} \delta(1-\rho)\, C(s)\, ds. \tag{11}$$

This is finite, since D is bounded. Hence, $\delta(1-\rho)\, C(t) \to 0$ as $t \to +\infty$. Since $\delta(1-\rho) > 0$, we have $C(t) \to 0$ as $t \to +\infty$. Using the same argument and noting that

$$C^* - C(0) = \int_0^{+\infty} \epsilon\, I_1(s) + \epsilon\, I_2(s) + \epsilon\, I_3(s) - \delta\, C(s)\, ds, \tag{12}$$

we also have $I_1(t)$, $I_2(t)$, $I_3(t) \to 0$ as $t \to +\infty$. Similar deduction can be used to show that $R(t) \to 0$ as $t \to +\infty$, using Equation (7). □

We may define the effective reproduction number associated with strain i by

$$\mathcal{R}_i^e(t) = \left(\frac{\lambda_1\alpha\beta_i}{\nu(\gamma_i+\epsilon)}\right)\frac{V(t)}{N_0}. \tag{13}$$

Compared to the basic reproduction number $\mathcal{R}_0$, the effective reproduction number $\mathcal{R}_i^e(t)$ denotes the average number of new infections associated with strain i, at time t, caused by a single infected individual, considering that in the population at this time, there are already some individuals who are no longer susceptible.

4. Existence of solution for the SDE Model

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$ be a complete probability space with filtration $\{\mathcal{F}_t\}_{t\geq 0}$ satisfying the usual conditions (i.e., it is increasing and right continuous while $\mathcal{F}_0$ contains all $\mathbb{P}$-null sets). Let $\mathbb{R}^8_+ = \{x_i > 0 : i = 1, 2, ..., 8\}$. Let $B(t)$ be a Brownian motion defined on the complete probability space Ω. Then, we have the following theorem showing that the stochastic system (9) has a unique non-negative global solution.

Theorem 4.1. *For any given initial value* $x_0 \in \mathbb{R}^8_+$, *there is a unique solution* $x(t)$ *of system (9) on* $t \geq 0$, *and the solution will remain in* $\mathbb{R}^8_+$ *with probability 1, namely,* $x(t) \in \mathbb{R}^8_+$ *for all* $t \geq 0$ *almost surely.*

Proof. Since the coefficients of system (9) satisfy the local Lipschitz condition, it implies that for any given initial value $x_0 \in \mathbb{R}^8_+$, there is a unique local solution $x(t)$ for every $t \in [0, \tau_e)$, where τ_e is the explosion time. To prove that the solution is global, we need to show that $\tau_e = \infty$. To do so, we let $s_0 \geq 1$ be sufficiently large so that all components of x_0 are contained in the interval $[1/s_0, s_0]$. For each integer $s \geq s_0$, we define the stopping time by

$$\tau_s = \inf\{t \in [0, \tau_e) : \text{at least one of } S, V, I_1, I_2, I_3, C, R, \text{ or } D \in (1/s, s)\}.$$

Clearly, τ_s is increasing as $s \to \infty$. Let $\tau_\infty = \lim_{s\to\infty} \tau_s$, then $\tau_e \geq \tau_\infty$ almost surely. If we can show that $\tau_\infty = \infty$ a.s., then $\tau_e = \infty$ and $x(t) \in \mathbb{R}^8_+$ a.s. for all $t \geq 0$.

Suppose $\tau_\infty < \infty$, then there exists $T > 0$ such that $\mathbb{P}\{\tau_\infty \leq T\} > \epsilon$ for all $\epsilon \in (0, 1)$. Hence, there is an integer $s_1 \geq s_0$ such that

$$\mathbb{P}\{\tau_s \leq T\} > \epsilon, \quad \text{for every } s \geq s_1. \tag{14}$$

Let a $\mathcal{C}^2-$ function $U : \mathbb{R}^8_+ \to \mathbb{R}^8_+$ be defined by

$$\begin{aligned} U(x) &= S - 1 - \ln(S) + V - 1 - \ln(V) + I_1 - 1 - \ln(I_1) \\ &+ I_2 - 1 - \ln(I_2) \\ &+ I_3 - 1 - \ln(I_3) + C - 1 - \ln(C) + R - 1 - \ln(R) \\ &+ D - 1 - \ln(D). \end{aligned} \tag{15}$$

Using Itô formula on Equation (15), we have

$$dU = LUdt + \sigma\frac{I_3}{N_0}dB - \sigma\frac{S}{N_0}dB,$$

where

$$\begin{aligned} LU = &\left(1-\frac{1}{S}\right)\left[-\alpha(\beta_1 I_1 + \beta_2 I_2 + \beta_3 I_3)\frac{S}{N_0} - \nu S + \lambda_1 V + \lambda_2 R\right] \\ &+\frac{1}{2}\sigma^2\frac{I_3^2}{N_0^2} + \left(1-\frac{1}{V}\right)\left[\nu S - \lambda_1 V\right] + \left(1-\frac{1}{I_1}\right) \\ &\left[\alpha\beta_1 I_1\frac{S}{N_0} - (\gamma_1+\epsilon)I_1\right] + \left(1-\frac{1}{I_2}\right) \\ &\left[\alpha\beta_2 I_2\frac{S}{N_0} - (\gamma_2+\epsilon)I_2\right] + \left(1-\frac{1}{I_3}\right) \\ &\left[\alpha\beta_3 I_3\frac{S}{N_0} - (\gamma_3+\epsilon)I_3\right] + \frac{1}{2}\sigma^2\frac{S^2}{N_0^2} + \left(1-\frac{1}{C}\right) \end{aligned}$$

$$\left[\epsilon(I_1+I_2+I_3)-\delta C\right]+\left(1-\frac{1}{R}\right)$$
$$\left[\gamma_1 I_1+\gamma_2 I_2+\gamma_3 I_3+\delta\rho C-\lambda_2 R\right]$$
$$+\left(1-\frac{1}{D}\right)\left[\delta(1-\rho)C\right]\leq \nu+\lambda_1+(\gamma_1+\gamma_2+\gamma_3)+3\epsilon$$
$$+\delta+\lambda_2+\frac{1}{2}\sigma^2\frac{I_3^2}{N_0^2}+\frac{1}{2}\sigma^2\frac{S^2}{N_0^2}+\alpha\beta_1\frac{I_1}{N_0}+\alpha\beta_2\frac{I_2}{N_0}$$
$$+\alpha\beta_3\frac{I_3}{N_0}\leq \nu+\lambda_1+(\gamma_1+\gamma_2+\gamma_3)+3\epsilon+\delta+\lambda_2$$
$$+\sigma^2\frac{N_0^2}{N_0^2}+\left[\alpha\beta_1+\alpha\beta_2+\alpha\beta_3\right]\frac{N_0}{N_0}\leq \nu+\lambda_1$$
$$+(\gamma_1+\gamma_2+\gamma_3)+3\epsilon+\delta+\lambda_2+\sigma^2+\alpha\beta_1+\alpha\beta_2+\alpha\beta_3$$
$$:=K.$$

Note that K is a positive constant independent of the variables S, V, I_1,I_2, I_3, C, R, and D, and time t.
Thus,

$$dU=Kdt+\sigma\frac{I_3}{N_0}dB-\sigma\frac{S}{N_0}dB. \tag{16}$$

Therefore, if $t_1\leq T$,

$$\int_0^{\tau_s\wedge t_1}dU\leq\int_0^{\tau_s\wedge t_1}Kdt+\int_0^{\tau_s\wedge t_1}\frac{\sigma}{N_0}(I_3-S)dB(t), \tag{17}$$

where $\tau_s\wedge t_1=\min\{\tau_s,t_1\}$. Taking expectations to both sides of (17), we obtain

$$\mathbb{E}\left[\int_0^{\tau_s\wedge t_1}dU\right]\leq\mathbb{E}\left[\int_0^{\tau_s\wedge t_1}Kdt\right]+\mathbb{E}\left[\int_0^{\tau_s\wedge t_1}\frac{\sigma}{N_0}(I_3-S)dB(t)\right]. \tag{18}$$

By properties of Itô integral, we have

$$\begin{aligned}\mathbb{E}U(F(\tau_s\wedge t_1),C(\tau_s\wedge t_1),\rho(\tau_s\wedge t_1))&\leq U(F(0),C(0),\rho(0))\\&+\mathbb{E}\left[\int_0^{\tau_s\wedge t_1}Kdt\right]\\&\leq U_0+KT<\infty,\end{aligned} \tag{19}$$

where $U_0=U(x_0)$. By Gronwall's inequality,

$$\mathbb{E}V(F(\tau_s\wedge t_1),C(\tau_s\wedge t_1),\rho(\tau_s\wedge t_1))\leq U_0e^{KT}<\infty. \tag{20}$$

Let $\Omega_s=\{\tau_s\leq T\}$ for any $s\geq s_1$. Then, by (14), we have $\mathbb{P}(\Omega_s)\geq\epsilon$. Note that for every $\omega\in\Omega_s$, there is at least one of $S(\tau_s,\omega)$, $V(\tau_s,\omega)$, $I_1(\tau_s,\omega)$, $I_2(\tau_s,\omega)$, $I_3(\tau_s,\omega)$, $C(\tau_s,\omega)$, $R(\tau_s,\omega)$, and $D(\tau_s,\omega)$ that is equal to either s or $1/s$. Consequently, $U(x(\tau_s,\omega))$ is no less than either

$$s-1-\ln(s)\text{ or }\frac{1}{s}-1-\ln\left(\frac{1}{s}\right)=\frac{1}{s}-1+\ln(s).$$

Thus,

$$U(x(\tau_s,\omega))\geq\left[s-1-\ln(s)\right]\wedge\left[\frac{1}{s}-1+\ln(s)\right].$$

TABLE 2 Values of the parameters used in the simulations.

Parameter	Value	Source
β_1	0.15094	Fitted
β_2	0.30187 (Day 130 onward)	Estimated [17]
β_3	0.90561 (Day 334 onward)	Estimated [18]
α	0.99964 (Day 0–Day 86)	Fitted
	0.83186 (Day 87–Day 165)	
	0.61940453 (Day 166–Day 366)	
	0.22039 (Day 367–Day 454)	
	0.26799 (Day 455 onward)	
ν	$270000/N_0$ (Day 87 onward)	Estimated [19]
λ_1	1/180, 1/365, 1/540 (Day 87 onward)	Simulated
λ_2	1/365	Estimated [20]
ϵ	0.03677	[11]
γ_i	1/12	[11]
δ	1/7	[11]
ρ	0.98	[21]

TABLE 3 Initial conditions used in the simulations.

Initial condition	Value	Source
$S(0)$	107,222,344	Estimated [11, 21, 22]
$I_1(0)$	17,494	Fitted
$I_2(0)$	4,523	Fitted
$I_3(0)$	3,825	Fitted
$C(0)$	26,677	[21]
$R(0)$	1,759,580	Estimated [11, 21]
$D(0)$	9,248	[21]

At Day 0, $V(0)=0$ and $N_0=109035343$, based on the Philippine Statistics Authority's 2020 estimated population of the country [22].

It then follows from Equation (14) and Equation (21) that

$$\begin{aligned}U_0e^{KT}&\geq\mathbb{E}[\mathbf{1}_{\Omega_s}(\omega)U(x(\tau_s,\omega))]\\&\geq\epsilon\left[s-1-\ln(s)\wedge\frac{1}{s}-1+\ln(s)\right],\end{aligned}$$

where $\mathbf{1}_{\Omega_s}$ is an indicator function of Ω_s. Letting $s\to\infty$, then we have

$$\infty>U_0e^{KT}\geq\infty,$$

which yields the contradiction. Therefore, we must have $\tau_\infty=\infty$, almost surely.□

5. Numerical simulations

Data on confirmed cases used in parameter calibration are from the COVID-19 Data Repository by the Center for Systems Science

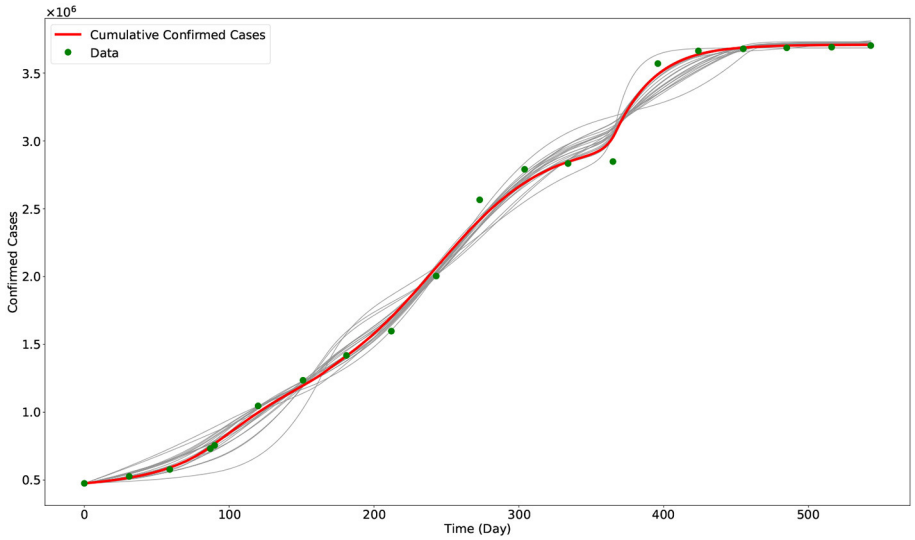


FIGURE 2
Output of the model with calibrated parameters compared with data. The gray curves represent the other outputs of the approximate Bayesian computation approach, while the red curve represents the best fit model output (i.e., using the parameters in Table 2). The green dots are the data from the COVID-19 data repository of JHU CSSE.

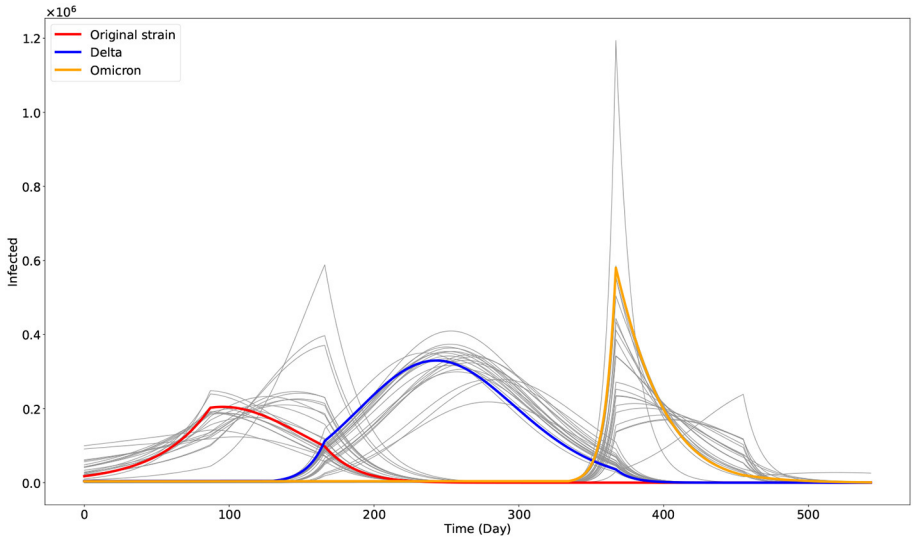


FIGURE 3
Plot of the three dominant COVID-19 variants in the Philippines as given by the model. The gray curves are the other outputs of the approximate Bayesian computation approach. The red, blue, and orange curves represent the best fit model output associated with the original strain (I_1), Delta (I_2), and Omicron (I_3) variants, respectively.

and Engineering (CSSE) at Johns Hopkins University (JHU) [16]. The data are publicly available and so ethical approval is not required. The data are from 1 January 2021 (Day 0) to 28 June 2022 (Day 543). Values of some parameters and initial conditions are taken or estimated from sources as indicated in Tables 2, 3. We note that the parameter α varies over time as controls implemented by the government also vary. Hence, we considered α as a piecewise constant function, as shown in Table 2, where the dates correspond to the noticeable changes in the control measures implemented by the government.

To obtain the fitted values for the parameters β_1 and α, and the initial conditions $I_1(0)$, $I_2(0)$, and $I_3(0)$, we minimized a non-linear least square function given by the sum of the square of the difference of the data and the model output. The optimization problem is solved using the approximate Bayesian computation approach combined with the Levenberg–Marquardt algorithm [23–25]. Visualization of the optimization result is given in Figure 2. In Figure 3, we plotted the evolution of the three variants based on the optimization result.

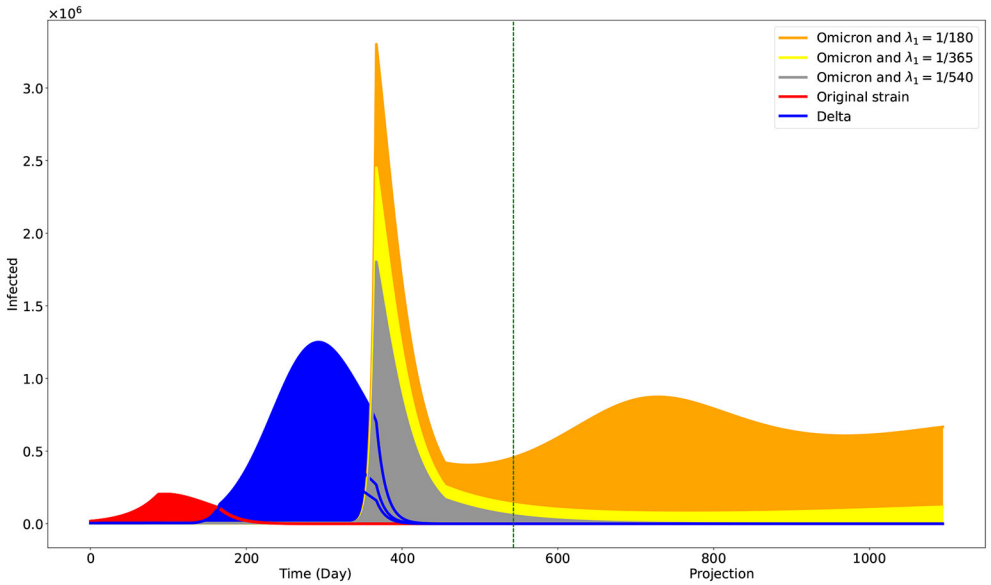


FIGURE 4
Projection on the effects of different levels of vaccine-induced immunity duration. The green vertical line is at Day 543, the end of the data used. One may interpret the case when $\lambda_1 = 1/180$ to be the case when individuals only opt to be fully vaccinated. The cases $\lambda_1 = 1/365$ and $\lambda_1 = 1/540$ may correspond to the cases when individuals are also getting first and second booster shots, respectively.

5.1. Deterministic simulations on waning vaccine-induced immunity

The Philippines started its vaccination campaign last 1 March 2021, with the two-dose Sinovac vaccine [26]. We estimated that it would need 4 weeks or 28 days to be fully protected from the vaccine [11], so we started the parameters ν and λ_1 by Day 87 (29 March 2021). It is estimated that the vaccine-induced immunity wanes after 6 months and that booster shots are recommended after that time interval [27]. Some people are unwilling to take booster shots due to vaccine hesitancy. Accounting for this social behavior in the model, we consider three different vaccine-induced immunity duration through the parameter λ_1 by setting it to $1/180$ (6 months). This accounts for the case when the population only takes full vaccination but no booster shots. On the contrary, we set $\lambda_1 = 1/365$ (12 months) when the people finished taking the first booster shot, while $\lambda_1 = 1/540$ (18 months) when they completed the second booster shot. We simulate up to 31 December 2023 (Day 1094). The result is shown in Figure 4.

5.2. Stochastic simulations on the dominant transmission rate

As shown in Figure 4, our simulation suggested that the original and Delta variants are to die out even with the minimum vaccine-induced immunity duration of 6 months. But it is not the case with regard to the Omicron variant. However, we acknowledge that much uncertainty can affect the reduced transmission rate $\alpha\beta_3$—for instance, the changing control measures of the government and the further evolution of the variant. Hence, we also want simulations incorporating noise on $\alpha\beta_3$. In Figure 5, we show the result of our stochastic simulations. The curves are for the confirmed compartment. The figure can be viewed as a 3×2 matrix where the rows correspond to the cases corresponding to the different vaccine-induced immunity duration, given by the value of λ_1. The columns show the results concerning different noise intensities, given by the value of σ. The stochastic simulations start after Day 543 and run up to 31 December 2023 (Day 1094). The numerical simulations are implemented using the Euler–Maruyama scheme [28].

6. Discussion

With our deterministic model, we have obtained parameters fitting observed data on confirmed cases in the Philippines. Our model has the added value that it could estimate the progression of the three main variants circulating the Philippines. Our simulations showed that the Omicron variant would be the dominant variant as we advance and the other variants die out (Figure 4). Moreover, it showed the significantly faster transmission of the Omicron variant compared with the other two, as reported [29] and validated by some studies [30, 31]. Our population is a closed model, and Theorem 3.1 showed that the Omicron variant would eventually die out. However, the time of its realization depends on the vaccine-induced immunity duration. With a duration of approximately 6 months, the simulation shows a noticeable presence of the Omicron variant by 31 December 2023. However, this presence is reduced significantly with vaccine-induced immunity durations of 12–18 months. We note that these vaccine-induced immunity durations can be achieved with booster shots in the population. Looking at the effective reproduction number given in (13), by Day 543 and vaccine-induced immunity duration of 6 months, we have $\mathcal{R}_1^e = 0.1731$, $\mathcal{R}_2^e = 0.3463$, and $\mathcal{R}_3^e = 1.0390$. By Day 5000,

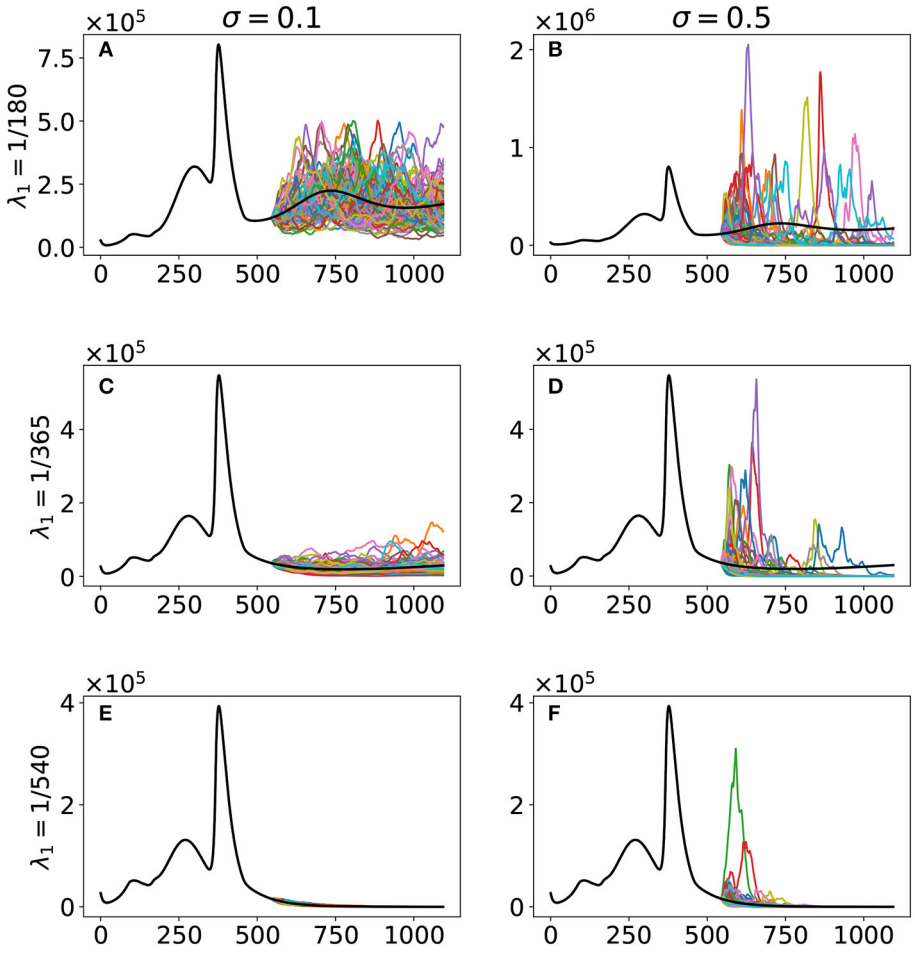


FIGURE 5
Stochastic simulations with respect to the vaccine-induced immunity duration λ_1 and noise intensity σ. The curves are for the confirmed compartment. The black curve is the deterministic model output, while the colored curves are the realizations of the 100 runs of the stochastic model. The cases are as follows: **(A)** $\lambda_1 = 1/180$, $\sigma = 0.1$; **(B)** $\lambda_1 = 1/180$, $\sigma = 0.5$; **(C)** $\lambda_1 = 1/365$, $\sigma = 0.1$; **(D)** $\lambda_1 = 1/365$, $\sigma = 0.5$; **(E)** $\lambda_1 = 1/540$, $\sigma = 0.1$; **(F)** $\lambda_1 = 1/540$, $\sigma = 0.5$.

we only have $\mathcal{R}_3^e = 0.9998$. However, with a vaccine-induced immunity duration of 12 months and at Day 543, we already have $\mathcal{R}_3^e = 0.5124$.

We acknowledge that a lot of changes can happen in the future. For instance, the ongoing evolution of the virus and the ever-changing control measures being implemented by the government. Hence, incorporating these uncertainties in our simulations can prove beneficial. Our stochastic simulations (Figure 5) reveal that uncertainties, represented as noise in our model, can significantly affect future outcomes when vaccine-induced immunity duration is only 6 months, as seen in the spread of the stochastic model output. An increase in vaccine-induced immunity duration means a decrease in the variability of our model output and hence better projection. The government can then use better projection to design more robust and refined intervention strategies to control the virus effectively.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://github.com/CSSEGISandData/COVID-19.

Author contributions

RC-a and JA conceptualized the paper and drafted the manuscript. HC, MR, and RC-a did the mathematical analysis. JM did the biological part. RC-a performed the simulations. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

This project was funded by the Philippines' Department of Science and Technology – Philippine Council for Health Research and Development, MSU – Iligan Institute of Technology's Office of the Vice-Chancellor for Research and Extension through the Premier Research Institute of Science and Mathematics, and the Caraga State University Mathematical, Statistical & Computing Research Center.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be ɔnstrued as a potential conflict of interest.

Publisher's note

References

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med*. (2020) 382:727–33. doi: 10.1056/NEJMoa2001017

2. World Health Organization. *Tracking SARS-CoV-2 Variants* (2022). Retrieved from: https://www.who.int/activities/tracking-SARS-CoV-2-variants

3. The World Bank. *Reducing Vaccine Hesitancy in the Philippines: Findings from a Survey Experiment* (2021). Retrieved from: https://thedocs.worldbank.org/en/doc/9b206c064482a4fbb880ee23d6081d52-0070062021/original/Vaccine-Hesitancy-World-Bank-Policy-Note-September-2021.pdf

4. World Health Organization. *Donors Making a Difference: Knocking Down Obstacles to COVID-19 Vaccination* (2022). Retrieved from: https://www.who.int/news-room/feature-stories/detail/donors-making-a-difference-knocking-down-obstacles-to-covid-19-vaccination

5. Arcede JP, Caga-anan RL, Mentuda CQ, Mammeri Y. Accounting for symptomatic and asymptomatic in a SEIR-type model of COVID-19. *Math Modell Nat Phenomena*. (2020) 15:34. doi: 10.1051/mmnp/2020021

6. Macalisang JM, Caay ML, Arcede JP, Caga-anan RL. Optimal control for a COVID-19 model accounting for symptomatic and asymptomatic. *Comput Math Biophys*. (2020) 8:168–79. doi: 10.1515/cmb-2020-0109

7. Bock W, Bornales JB, Burgard JP, Babiera JE, Caga-anan RL, Carmen DJS, et al. Testing, social distancing and age specific quarantine for COVID-19: case studies in Iligan City and Cagayan de Oro City, Philippines. In: *AIP Conference Proceedings*. AIP Publishing LLC (2020). doi: 10.1063/5.0029818

8. Arcede JP, Caga-anan RL, Mammeri Y, Namoco RA, Gonzales ICA, Lachica ZP, et al. A modeling strategy for novel pandemics using monitoring data: the case of early COVID-19 pandemic in Northern Mindanao, Philippines. *SciEnggJ*. (2022) 15, 35-46.

9. Arcede JP, Basañez RC, Mammeri Y. Hybrid modeling of COVID-19 spatial propagation over an Island Country. In: *Advances in Computational Modeling and Simulation*. Singapore: Springer (2022). p. 75–83.

10. Estadilla CDS, Uyheng J, de Lara-Tuprio EP, Teng TR, Macalalag JMR, Estuar MRJE. Impact of vaccine supplies and delays on optimal control of the COVID-19 pandemic: mapping interventions for the Philippines. *Infect Dis Poverty*. (2021) 10:46–59. doi: 10.1186/s40249-021-00886-5

11. Caga-anan RL, Raza MN, Labrador GSG, Metillo EB, del Castillo P, Mammeri Y. Effect of vaccination to COVID-19 disease progression and herd immunity. *Comput Math Biophys*. (2021) 9:262–72. doi: 10.1515/cmb-2020-0127

12. Cabanilla KI, Enriquez EAT, Velasco AC, Mendoza VMP, Mendoza R. Optimal selection of COVID-19 vaccination sites in the Philippines at the municipal level. *PeerJ*. (2022) 10:e14151. doi: 10.7717/peerj.14151

13. Buhat CAH, Duero JCC, Felix EFO, Rabajante JF, Mamplata JB. Optimal allocation of COVID-19 test kits among accredited testing centers in the Philippines. *J Healthcare Inform Res*. (2021) 5:54–69. doi: 10.1007/s41666-020-00081-5

14. Diekmann O, Heesterbeek JAP, Metz JA. On the definition and the computation of the basic reproduction ratio R 0 in models for infectious diseases in heterogeneous populations. *J Math Biol*. (1990) 28:365–82. doi: 10.1007/BF00178324

15. Van den Driessche P, Watmough J. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Math Biosci*. (2002) 180:29–48. doi: 10.1016/S0025-5564(02)00108-6

16. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Inf Dis*. (2020) 20:533–4. doi: 10.1016/S1473-3099(20)30120-1

17. American Society for Microbiology. *How Dangerous Is the Delta Variant (B.1.617.2)?* (2021). Retrieved from: https://asm.org/Articles/2021/July/How-Dangerous-is-the-Delta-Variant-B-1-617-2

18. Virginia Department of Health. *Variants of the Virus that Causes COVID-19*. Retrieved from: https://www.vdh.virginia.gov/coronavirus/get-the-facts/variants-of-covid

19. DOH. *National COVID-19 Vaccination Dashboard* (2022). Retrieved from: https://doh.gov.ph/covid19-vaccination-dashboard

20. Shrestha NK, Burke PC, Nowacki AS, Terpeluk P, Gordon SM. Necessity of coronavirus disease 2019 (covid-19) vaccination in persons who have already had COVID-19. *Clin Infect Dis*. (2022) 75:e662–71. doi: 10.1093/cid/ciac022

21. DOH. *COVID-19 Bulletin # 293* (2022). Retrieved from: https://doh.gov.ph/covid19casebulletin293

22. Philippine Statistics Authority. Retrieved from: https://psa.gov.ph/content/2020-census-population-and-housing-2020-cph-population-counts-declared-official-president

23. Csilléry K, Blum MG, Gaggiotti OE, Francois O. Approximate Bayesian computation (ABC) in practice. *Trends Ecol Evol*. (2010) 25:410–8. doi: 10.1016/j.tree.2010.04.001

24. Levenberg K. A method for the solution of certain non-linear problems in least squares. *Q Appl Math*. (1944) 2:164–8. doi: 10.1090/qam/10666

25. Marquardt DW. An algorithm for least-squares estimation of nonlinear parameters. *J Soc Indus Appl Math*. (1963) 11:431–41. doi: 10.1137/0111030

26. Parrocha A. *PH Kick-Starts COVID-19 Vaccination Drive with CoronaVac*. Philippine News Agency (2021). Retrieved from: https://www.pna.gov.ph/articles/1132127

27. World Health Organization. *Interim Statement on the Use of Additional Booster Doses of Emergency Use Listed mRNA Vaccines Against COVID-19* (2022). Retrieved from: https://www.who.int/news/item/17-05-2022-interim-statement-on-the-use-of-additional-booster-doses-of-emergency-use-listed-mrna-vaccines-against-covid-19

28. Kloeden PE, Platen E. Stochastic differential equations. In: *Numerical Solution of Stochastic Differential Equations*. Berlin; Heidelberg: Springer (1992). p. 103–60 doi: 10.1007/978-3-662-12616-5_4

29. World Health Organization. *One Year Since the Emergence of COVID-19 Virus Variant Omicron* (2022). Retrieved from: https://www.who.int/news-room/feature-stories/detail/one-year-since-the-emergence-of-omicron

30. Ren SY, Wang WB, Gao RD, Zhou AM. Omicron variant (B. 1.1. 529) of SARS-CoV-2: mutation, infectivity, transmission, and vaccine resistance. *World J Clin Cases*. (2022) 10:1. doi: 10.12998/wjcc.v10.i1.1

31. Vitiello A, Ferrara F, Auti AM, Di Domenico M, Boccellino M. Advances in the omicron variant development. *J Intern Med*. (2022) 292:81–90. doi: 10.1111/joim.13478

frontiers | Frontiers in Applied Mathematics and Statistics

TYPE Original Research
PUBLISHED 31 March 2023
DOI 10.3389/fams.2023.1000785

OPEN ACCESS

EDITED BY
Mohd Hafiz Mohd,
University of Science Malaysia (USM), Malaysia

REVIEWED BY
Hirohide Haga,
Doshisha University, Japan
Mohd Shareduwan Mohd Kasihmuddin,
University of Science Malaysia (USM), Malaysia

*CORRESPONDENCE
Meritxell Vinyals
✉ meritxell.vinyals@inrae.fr
Patrick Taillandier
✉ patrick.taillandier@inrae.fr

SPECIALTY SECTION
This article was submitted to Mathematics of Computation and Data Science, a section of the journal Frontiers in Applied Mathematics and Statistics

RECEIVED 22 July 2022
ACCEPTED 13 March 2023
PUBLISHED 31 March 2023

CITATION
Vinyals M, Sabbadin R, Couture S, Sadou L, Thomopoulos R, Chapuis K, Lesquoy B and Taillandier P (2023) Toward AI-designed innovation diffusion policies using agent-based simulations and reinforcement learning: The case of digital tool adoption in agriculture. *Front. Appl. Math. Stat.* 9:1000785. doi: 10.3389/fams.2023.1000785

Toward AI-designed innovation diffusion policies using agent-based simulations and reinforcement learning: The case of digital tool adoption in agriculture

Meritxell Vinyals[1]*, Regis Sabbadin[1], Stéphane Couture[1], Loïc Sadou[1], Rallou Thomopoulos[2], Kevin Chapuis[3], Baptiste Lesquoy[4,5] and Patrick Taillandier[1,4,5]*

[1]UR 875, MIAT, INRAE, Toulouse University, Castanet Tolosan, France, [2]UMR 1208, IATE, Univ Montpellier, INRAE, Institut Agro, Montpellier, France, [3]UMR 228, ESPACE-DEV, IRD, Montpellier, France, [4]UMI 209, UMMISCO, IRD, Sorbonne Université, Bondy, France, [5]LMI ACROSS, Thuyloi University, Hanoi, Vietnam

In this paper, we tackle innovation diffusion from the perspective of an institution which aims to encourage the adoption of a new product (i.e., an innovation) with mostly social rather than individual benefits. Designing such innovation adoption policies is a very challenging task because of the difficulty to quantify and predict its effect on the behaviors of non-adopters and the exponential size of the space of possible policies. To solve these issues, we propose an approach that uses agent-based modeling to simulate in a credible way the behaviors of possible adopters and (deep) reinforcement learning to efficiently explore the policy search space. An application of our approach is presented for the question of the use of digital technologies in agriculture. Empirical results on this case study validate our scheme and show the potential of our approach to learn effective innovation diffusion policies.

KEYWORDS

innovation diffusion, policy design, reinforcement learning, agent-based simulation, deep reinforcement learning, digital agriculture

1. Introduction

Many areas such as agriculture have been transformed by the arrival of new innovations. These transformations are sometimes desired by institutions that wish to promote for instance more environmentally friendly models. In this paper, we propose to tackle this issue from the perspective of an institution which aims to encourage the adoption of some innovation. In particular, we focus on domains in which individuals are reluctant, for a variety of possible reasons, to adopt the innovation even when it can benefit them: misinformation, distrust of the hidden agenda of the institutions promoting it, lack of skills

to use it, etc. In this situation the institution may make use of different policy actions (information, financial aid, training,...) in order to overcome these barriers hindering the widespread adoption of the innovation. However, designing such innovation adoption policies is a very challenging task because of: (i) the difficulty to quantify and predict the effect of policies on the behaviors of non-adopters (e.g., the effect of an advertising campaign can be widespread over time and difficult to differentiate from the effect of other similar actions); and (ii) the exponential size of the policy search space. In more detail, when designing an innovation diffusion policy the institution faces a budget-constrained sequential decision making problem in which, at each time step, it needs to decide if it launches new (parallel) actions on several areas of interest as well as the parameters that accompany each of these actions. Therefore, it is not only which actions but when, with which combination and in which parameterization. Moreover, to perform such decisions the institution has only access to some aggregated indicators at global level (e.g., the number of adopters) and none at individual level, resulting in a very partial view of the state of the environment.

Against this background, we propose to study the problem of designing innovation diffusion policies using agent-based simulations and reinforcement learning.

On the one hand, a computational simulation environment is necessary given the limited opportunities to experiment with policy-making in the real-world. In this context, agent-based simulation has proven to be an effective tool to study the complex social dynamics that emerge in the diffusion of innovation among an heterogeneous population of potential adopters [1]. In particular, agent-based modeling addresses the limitations of aggregate models [2] by explicitly representing individuals, their social interactions, and their decision-making processes. The difficulty that remains is how to define credible behaviors for the agents while keeping the computational cost of the simulations affordable. In particular, in innovation diffusion we need to model the human decision-making key-factors deciding on innovation adoption. We tackle this problem by building an agent-based model of the innovation diffusion process based on the theory of planned behavior [3] to describe the adoption of a new behavior.

On the other hand, even when the institution is expected to be able to adapt the public policy in place to the context, most current works are content to represent it as a fixed variable or a set of scenarios to explore [4]. Instead, in this paper, the institution is modeled as an intelligent agent which learns, by reinforcement learning (RL), how to adapt the public policy over time. In particular, and as it is common in the literature when dealing with highly complex partially-observable environments, we based our approach on *deep* reinforcement learning, where deep stands for an artificial deep neural network (NN) that is used to approximate the policy function. Recent works have already shown the capacity of deep RL to automatically learn public policies, but on markedly different domains, e.g., on the problem of designing taxation [5], pandemic response [6] and market-price intervention [7] policies. However, to the best of our knowledge, no previous work has used (deep) RL to address the problem of designing public policies that maximize the number of adopters of an innovation, i.e., the so-called innovation diffusion policy design problem that we formalize in this paper. As we analyze in this work, it turns out that the action space of this problem is particularly complex due to the fact that: (i) an action is composed of a set of sub-actions, corresponding to different types of initiatives that the institution launches in parallel; (ii) each sub-action is parameterized by a set of (real-valued) parameters that accompany that action; (iii) actions in a given time are constrained by the available budget, shared among all sub-actions. These characteristics make standard deep RL methods not directly applicable to the innovation diffusion policy design problem. The problem can be cast as a Constrained Markov Decision Process (CMDP) and optimized by one of the general-purpose approaches which have been proposed to solve CMDPs [8]. However, such approaches typically negatively affect the performance and scalability (or both) of the learning compared with the non-constrained case. Given this, in this paper we instead opted for a specific architecture that is able to exploit the structure of the particular action space of this problem. More precisely, we propose a NN policy architecture in which the policy learns the budget allocation among the multiple action types along with the values of the parameters of those actions. All in all, our approach allows for AI-designed policies for innovation diffusion.

We illustrate the use of our approach on a particular application in the framework of digital technologies in agriculture and more particularly, on the adoption of communicating water meters by farmers in the Louts region (South-West of France). Nowadays, farmers in this area mainly use mechanical meters which poorly estimate water consumption due to a low accuracy. This is an advantage for farmers, as they are less likely to be overcharged if they exceed the allocated quota. Nevertheless, this over-consumption is an important issue in this region where the water level of the rivers tends to decrease every year due to climate change, which has major consequences for the local ecosystem. For this reason, the Ministry of the Environment has required a periodic refurbishment of the metering system every 9 years. The institution in charge of managing water distribution in this area is counting on this regulation to install its new communicating meters. These new meters are more accurate and allow for real-time monitoring of each farmer's consumption and thus better manage water use. However, the institution is having difficulty convincing farmers to install this device because they perceive it negatively. This obstacle is closely linked to the farmers' distrust of the institution. A large part of the farmers think that the new meter does not bring them anything and that it is only useful to the institution. Thus, the institution is now questioning the policy to be implemented in order to encourage the adoption of new meters.

Our main contributions in this article are the following:

- We propose an AI framework for the problem of designing effective innovation diffusion policies that combines agent-based simulations with (deep) reinforcement learning techniques;
- Our agent-based model for innovation diffusion combines for the first time three existing models ([4] for the theory of planned behavior to model the adoption of

the innovation, Deffuant et al. [9] and Deffuant et al. [10] for the bounded confidence model to represent the social influence among individuals and Sadou et al. [11] for the integration of several topics to build an opinion) with the objective to simulate the behavior of adopters in a credible way, while keeping the computational cost of simulations affordable;
- We model the institution as an intelligent agent that learns by (deep) reinforcement learning how to adapt the public policy in place to the particular observed context. The reinforcement learning mechanism of this institution agent extends previous approaches to deal with non mutually-exclusive parameterizable actions and budget constraints;
- We illustrate this generic model on the particular application of the adoption of digital tools in agriculture. We motivate this particular application with a real use case of the adoption of communicating meters by farmers in the Louts region (South-West of France);
- We provide a first validation of our approach showing, using simulations, how our institution agent can learn effective policies that promote the adoption of an innovation for this particular application.

The rest of this paper is structured as follows. Section 2 reviews the related work. Section 3 introduces our agent-based model of innovation diffusion, while Section 4 instantiates this model on the adoption of digital tools in agriculture. Section 5 details the reinforcement learning method used to search for an optimal policy and Section 6 presents our experimental validation. Finally, Section 7 draws conclusions and sets paths for future research.

2. Related work

Concerning the modeling of the innovation adoption process, many studies have tackled this topic with most of the models building on the work of Rogers [12] on the diffusion of innovations. Agent-based modeling is becoming increasingly popular for the study of this type of process [1], each agent representing an individual that can influence the others on their adoption of the innovation. As discussed in Kiesling et al. [1], aggregate models such as the classical Bass model [2] do not explicitly account for consumer heterogeneity and the complex dynamics of social processes involved in innovation diffusion. Agent-based modeling addresses this limitation by explicitly representing individuals, their social interactions, and their decision-making processes. Zhang and Vorobeychik [13] proposed a critical review of these agent-based models. In particular, they proposed to categorize these models based on how the models represent the decision to adopt. Among these categories, we can distinguish cognitive agent models that are closest to our concerns: they aim to explicitly represent how individuals influence each other in cognitive and psychological terms. A particularly popular model in this category is the *relative agreement* model of Deffuant et al. [9], which focuses on the notion of opinion about an innovation. The individual's opinion and uncertainties are represented by numerical values that evolve during interpersonal interactions. Other models are more interested in the adoption process as such: how an individual will decide whether or not to adopt an innovation. In this context, a classical approach is to base the decision of agents on the theory of planned behavior (TPB) [3]. This theory states that the intention to perform a behavior is a reliable predictor of the implementation of that behavior. The intention is derived from 3 factors: attitude, subjective social norm, and perceived behavioral control (PBC). The attitude represents the opinion that an individual has about the behavior. The subjective norm is the individual's perception of the adoption opinion of her/his social network. Finally, the PBC is the capacity felt by the individual to adopt the behavior (in terms of cost, time, skills...). A representative model of this use is that proposed by Bourceret et al. [14] concerning the adoption of more environmentally friendly agricultural practices. They propose a simple model based on the work of Beedell and Rehman [15] for the calculation of the intention from the attitude, the social norm and the PBC. In this model, the interaction between the agents is indirect through the social norm, there is no direct exchange between them. On the contrary, Sadou et al. [11] propose a more complex model, also based on the PBC, in which the agents try to convince each other explicitly through the exchange of arguments. In addition, this model introduces heterogeneity between agents by integrating the notion of point of view on different subjects (economy, environment...) which is used to compute the attitude of agents. While this model offers a powerful way of representing the change of opinion of agents, it requires a lot of data (arguments) and can be computationally heavy.

The model proposed is based on the three previous models: the model of Bourceret et al. [4] for its use of TPB for the decision making regarding the adoption of the innovation, the model of Deffuant et al. [9] and more particularly on the bounded confidence model [10] for the representation of the social influence between people, and finally the model of Sadou et al. [11] for the integration of several topics for the building of opinion.

Concerning the issue of governance representation in agent-based models, Bourceret et al. [4] have recently conducted a systematic review of the literature for socio-ecological issues. This review shows that in most works, governance is represented in the form of variables and not in the form of an agent: thus, if governance impacts the other agents, very few models take into account the fact that governance can be impacted by the other agents and in particular adapt the policy implemented according to the context. Generally, governance is just studied as a set of scenarios or parameters to be explored, which is not representative of most real contexts where the governance is able to adapt its policies according to the setting. Contrary to works like [4], we propose to represent it as an agent able to learn, by reinforcement learning (RL), how to adapt the policy to the context. Given the large body of work on reinforcement learning, on the remaining of this section we focus on the most relevant areas for our work, namely: deep RL for policy design and RL advances to deal with complex (e.g., parametrized, combinatorial, constrained, ..) action spaces.

Deep RL for policy design. Recently, some works [5–7] have also applied deep reinforcement learning to the problem of designing public policies. In Danassis et al. [7] a Deep RL policymaker agent adjusts the prices in a production market (e.g., the common-fishery market) to consider multiple objectives, including sustainability and resource wastefulness. Nevertheless, the output of this model is just a vector of continuous action values (i.e., corresponding to the price of each good in the next market round) and they do not consider any constraint on the budget of the policymaker (i.e., the cost of the policymaker intervention, defined as the difference between RL computed prices and the traditional competitive market prices, is not bounded). Zheng et al. [5] and Trott et al. [6] in the context of the so-called AI Economist frameworks applied Deep RL to the problem of designing taxation and pandemic response policies, respectively. However, in both works the policy outputs a single action discretized in intervals (so their setting ends up being a basic discrete action space) and do not consider any constraint on the cost of policy actions.

Parametrized action spaces. Several frameworks and algorithms [16, 17] have been proposed to deal with parametrized action spaces, in which the policy requires specified parameters associated to (discrete) action values. However, those works consider mutually-exclusive actions and the solution is typically built through a two-level decision making in which first the action to be applied is selected and, second, the parameters are defined for this action. Here we can not apply this type of approach since in our problem the institution can launch any combination of actions in parallel, i.e., at the same time step.

Combinatorial action spaces. Other works [18, 19] focused on action spaces where each action is a set of multiple interdependent sub-actions (i.e., non-mutually exclusive). However, in such works the complexity of the action space emerge from considering a large number of actions leading to an exponential combinatorial action space. Instead, in innovation diffusion policies the number of types of actions is typically small and the complexity rather emerges from the fact that actions are parametrized by a set of continuous parameters along with a budget constraint that creates interdependencies among them.

Bounded action spaces. Bounded action scenarios in discrete domains are typically addressed by action masking [20] whereas in continuous domains they are typically addressed by bounding the corresponding distributions [21]. However, given that the budget constraint is defined over every types of actions (creating interdependencies among them) we can not use individual bounding techniques here[1].

Constrained action spaces. Given the budget constraint, our work is also related to constrained policy learning. In Liu et al. [8], the problem of learning with constraints is modeled as a Constrained Markov Decision Process (CMDP). Existing approaches to solve CMDPs include enhancing: (i) the NN with an extra layer that projects actions onto a feasible space [22, 23] or (ii) the policy gradient algorithm itself [24–26]. However, given the general-purpose characteristic of these approaches (i.e., they can deal with any kind of constraint), they typically complexify the learning process, affecting its performance/scalability. Thus, even though the innovation diffusion policy design problem can be cast as a CMDP, in this work we take advantage of the fact that the budget constraint is explicit and known by the learner to exploit its structure and represent it in the policy network in a more expressive form (i.e., in our approach the policy network directly outputs the budget distribution among the multiple action types and once this is known we can act as in a bounded action space).

All in all, to the best of our knowledge, innovation diffusion policy design has not been tackled in the machine learning literature. Moreover, the complex action space of this problem, composed of multiple continuous parametrizable actions constrained by a common budget, argues for novel architectures that can deal with this action space efficiently.

TABLE 1 State variables of an *Individual* agent *i*.

State variable	Data type	Description
$social\_network^i$	List of *individual* agents-static	List of *individual* agents with whom *i* is in contact
$P^i_{interact}$	Float [0,1]-static	Probability of interaction with an *individual* agent on 1 day
$W^i_{attitude}$	Float [0,1]-static	Importance of the attitude in the intention computation
W^i_{social}	Float [0,1]-static	Importance of the social norm in the intention computation
W^i_{pbc}	Float [0,1]-static	Importance of the PBC in the intention computation
Ω^i	Float [0,1]-static	Adoption threshold
W^i_k	Float [0,1]-static	Importance of topic k
W^i_c	Float [0,1]-static	Importance of constraint c
$Adoption^i(t)$	Boolean-dynamic	Has the agent adopted the innovation at year $t \in \{1, .., H\}$?
μ^i	Float [0,1]-static	Speed of opinion convergence
d^i	Float [0,1]-static	Maximal opinion difference accepted for convergence
$I^i(t)$	Float [0,1]-dynamic	Current intention value
$A^i(t)$	Float [0,1]-dynamic	Current attitude value
$SN^i(t)$	Float [0,1]-dynamic	Current social norm value
$PBC^i(t)$	Float [0,1]-dynamic	Current PBC value
$Op^i_k(t)$	Float [0,1]-dynamic	Opinion on the topic k
$Skill^i_c(t)$	Float [0,1]-dynamic	Agent's capacity regarding a constraint c
$Support^i_k(t)$	Float [0,1]-dynamic	Value of the temporary support currently in place in relation to topic k
$Support^i_c(t)$	Float [0,1]-dynamic	Value of the temporary support currently in place in relation to constraint c

1 In our approach we use bounding to restrict action parameters to their feasible interval (which is independent of the state).

3. Agent-based model of innovation diffusion

This section presents the model following the Overview, Design concepts and Details (ODD) protocol [27].

3.1. Overview

3.1.1. Purpose and patterns

The model aims to assess the impact of the governance's policy regarding the adoption of an innovation. We evaluate our model by its ability to reproduce two patterns. The first one concerns the impact of interpersonal relation in the innovation diffusion process. Indeed, as stated in many works such as Rogers [12], the process of diffusion of innovation is partly due to interactions between people. The second concerns the capacity of institutions to promote the adoption of an innovation, as public policies can play an important role in the diffusion process [28].

3.1.2. Entities, state variables, and scales

Two types of entities are represented in the model: the possible adopters (*Individual* agents) and the *Institution* (unique agent). Tables 1, 2, respectively, present the state variables of the *Individual* and *Institution* entities.

Let H be the simulation length (total number of steps), then three time frames are taken into consideration (see Figure 1):

- Time frame of the interactions between the *Individual* agents (T^{ind})
- Time frame of the institution's action implementation (T^{inst})
- Time frame of the updating of the institution's budget (T^{budget})

There is no explicit representation of space.

3.1.3. Process overview and scheduling

Algorithm 1 details the pseudocode of the model. During the *Individual* agent interaction, each *Individual* agent i first has the probability $P^i_{interact}$ of interacting with another individual agent in $social\_network^i$ (randomly chosen) (lines 3–5), which will result in a potential convergence of their opinion on a topic (lines 6–7). In case of convergence, the *Individual* agents update their intention to adopt (line 8) and decide whether or not they wish to adopt the innovation (line 9). The details of the computation of the intention of the *Individual* agents and the interactions between them are given in Section 3.3.2. The agent adopts the innovation if the intention of the agent at time t, $I^i(t)$, is higher than the adoption threshold Ω^i:

$$\text{if } I^i(t) > \Omega^i\text{: } Adoption^i(t) = true \tag{1}$$

It is possible to consider two cases depending on the innovation: either an *Individual* agent cannot go back (once adopted, it keeps the innovation), or it can decide not to keep the innovation.

During the institution's action implementation time frame (line 11–13), first the supports of all agents will be set to *null*, then the *Institution* agent will choose which actions to implement. The actions can act directly on the *Individual* agents by permanently changing their opinion on a topic or their skills to handle a constraint, or they can be temporary (only during the period when the action is active). In the first case, the action will directly modify the opinion $Op^i_k(t)$ or skill $Skill^i_c(t)$ values of the agents at time t. In the second case, it will modify the support on a topic, $Support^i_k(t)$, or on a constraint, $Support^i_c(t)$ of the agents. The actions that the institution can implement depend on the field of application. As such, we present in Section 4.2 the actions of the institution for the case of the adoption of communicating water meters in agriculture.

```
while time < End_time do
  for all Individual agents i do
    if random(0.0, 1.0) ≤ P^i_interact then
      i' ← one_of(Individual agent in social_network^i)
      topic ← one_of(topics)
      if |Op^i_topic − Op^i'_topic| < threshold d^i then
        convergence of opinion on topic (Equations 9,
        10)
        updating of intention for i and i' (Equation
        2)
        updating of the adoption status for i and i'
        (Equation 1)
      end if
      if time modulo (6 months) = 0 then
        actions of the Institution agent
      end if
      if time modulo (1 year) = 0 then
        updating of the Institution agent's Budget
      end if
    end if
  end for
  time ← time + 1 day
end while
```

Algorithm 1. Pseudocode of the agent-based model of innovation diffusion.

Finally, during the updating of the institution's budget time frame (line 14–16), the agent *Institution* receives a new budget $Budget_{year}$ which is added to its current budget:

$$Budget(t) \leftarrow Budget(t) + Budget_{year}.$$

TABLE 2 State variables of the *Institution* agent at time step t.

State variable	Data type	Description
$Budget(t)$	Float-dynamic	Budget available to implement public policies
$Adopters(t)$	Float [0,1]-dynamic	Fraction of adopters
$Steps(t)$	Int-dynamic	Number of remaining decisions steps left to implement institutional actions (e.g., $T^{inst} - t$)

The computational complexity of the process described above for our agent-based model is linear in the number of agents.

3.2. Design concepts

3.2.1. Basic principles

The model is based on the use of the theory of planned behavior [3] to describe the adoption process of agents.

3.2.2. Interaction

Individual agents can interact directly with each other to try to influence each other. Also, *Individual* agents will interact indirectly through the social norm: the choice to adopt or not depends on the number of adopters in their social network.

The *Institution* agent will implement actions that will impact the *Individual* agents and in particular the opinion they have on the innovation with regard to different topics (at stake in the attitude computation), and also lift constraints to adoption (at stake in the PBC computation).

Also, there is an indirect interaction between the *Individual* and the *Institution* agents: the number of adopters can have an impact on the institution's decisions in terms of which action to implement.

3.2.3. Stochasticity

Apart from the initialization of the model which may involve stochasticity, an important element of stochasticity concerns the interactions between *Individual* agents: an *Individual* agent has a certain probability to choose to interact with another agent. The choice of the *Individual* agent with whom it interacts and the topic on which it wishes to discuss are chosen randomly.

Also, the choice of actions implemented by the *Institution* agent and the scope of these actions may involve stochasticity.

3.3. Details

3.3.1. Initialization

The initialization of the model includes the generation of the population of *Individual* agents: it is in particular a question of giving them values for all their attributes (linked to TPB, to the topics, to their social network...). Similarly, the *Institution* agent must be created and its attributes initialized. Section 4 gives an example of initialization of the model.

3.3.2. Submodels

3.3.2.1. Computation of the Intention

The intention of an *Individual* agent i is computed as follow:

$$I^i(t) \leftarrow W^i_{attitude} \times A^i(t) + W^i_{social} \times SN^i(t) + W^i_{pbc} \times PBC^i(t) \quad (2)$$

With:

$$W^i_{attitude} + W^i_{social} + W^i_{pbc} = 1 \quad (3)$$

The attitude A^i of an agent i considering the policy of the *Institution* agent is computed as follows:

$$A^i \leftarrow \sum_{k \in K} Min[1.0, Op^i_k(t) + Support^i_k(t)] \times W^i_k \quad (4)$$

With, K the set of adoption topics considered, and:

$$\sum_{k \in K} W^i_k = 1 \quad (5)$$

The social norm SN^i of an agent i is computed as follows:

$$SN^i(t) \leftarrow \frac{m^i(t)}{n^i} \quad (6)$$

where, $m^i(t)$ is the number of *Individual* agents of the social network $social\_network^i$ who have adopted at year t ($|\{i' \in social\_network^i, Adoption^{i'}(t) = true\}|$), and n^i is the number of *Individual* agents in $social\_network^i$.

The attitude $PBC^i(t)$ of an agent i at year t is computed as follows:

$$PBC^i(t) \leftarrow \sum_{c \in C} Min[1.0, Skill^i_c(t) + Support^i_c(t)] \times W^i_c \quad (7)$$

With, C the set of adoption constraints considered, and:

$$\sum_{c \in C} W^i_c = 1 \quad (8)$$

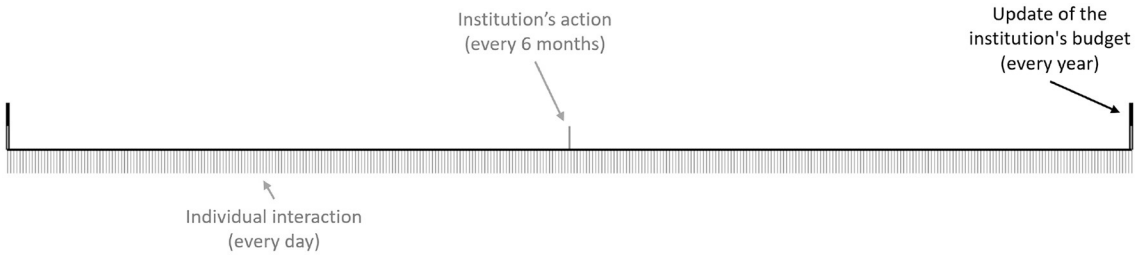


FIGURE 1
Time frames considered: every day, interaction of individuals; every 6 month's, implementation of the actions of the institution; every year, updating of the institution's budget.

3.3.2.2. Interaction between Individual agents

To manage interactions between *Individual* agents, we use the classic bounded confidence model [10]: when two *Individual* agents i and i' meet and begin a discussion on a topic k, they adjust their opinions on k as long as their difference of opinion on k is below a given threshold d^i. More formally, the opinions on k of i, i' are modified at time $t+1$ if $|Op_k^i(t) - Op_k^{i'}(t)| < d^i$ as follows:

$$Op_k^i(t+1) \leftarrow Op_k^i(t) + \mu \times \left(Op_k^{i'}(t) - Op_k^i(t)\right) \quad (9)$$

$$Op_k^{i'}(t+1) \leftarrow Op_k^{i'}(t) + \mu \times \left(Op_k^i(t) - Op_k^{i'}(t)\right) \quad (10)$$

4. Application for adoption of communicating water meters by farmers

The model presented in the previous section is intended to be completely generic and adaptable to any type of innovation[2]. In this section we instantiate this generic model for the particular application of the adoption of communicating water meters by farmers in South-West France.

4.1. Context

During the summer period, problems of water availability are common in some regions of France, particularly in the southern part of the country. These periods of drought have a direct impact on the daily life of irrigating farmers who must closely monitor their water consumption.

In a context of better resource management, various studies have highlighted the advantages that communicating meters could represent. Thus, on the Louts river (South-West of France), the Compagnie d'Aménagement des Coteaux de Gascogne (CACG), which is in charge of water distribution in this area, is proposing to irrigating farmers new communicating water meters to replace the aging mechanical meters. This is encouraged by the Ministry of the Environment, which requires the metering system to be replaced every 9 years.

Communicating water meters offer advantages over mechanical meters: they are more accurate and, above all, they allow for remote reading of consumption in real time. However, despite these advantages, CACG is having difficulty convincing farmers to install this device because, in general, they have a negative perception of it [29].

The questions then arise as to whether communicating meters will be adopted by farmers, what the impacts of the different information circulating on these devices are, and whether it is possible to implement public policies to promote a virtuous impact of these technologies. Answering these questions requires studying the social dynamics that lead to the adoption (or not) of an innovation within a population.

2 The model in Bourceret et al. [14] can be seen as a particular application of our model with low-input agricultural practices as innovation and with the following parameters: no interaction between individuals ($P^i_{interact}$ =0), 2 topics (Economy and Environment) and one constraint related to farmers' skills.

4.2. Specification of the model and description of the institution's actions

We have initialized the generic model presented in Section 3 for the case of the adoption of communicating water meters. In this model, *Individual* agents represent farmers. Based on the work of Sadou et al. [30] who have analyzed the arguments used by the stakeholders regarding communicating water meters, we identified 3 major topics: economy, environment and farm management (ease of management).

Concerning the adoption constraint, we chose to use the same type as Bourceret et al. [14], i.e., the technical skill.

We have integrated 3 types of actions for the institution: training, financial aid and environmental awareness. The actions are defined as follows:

- *Training*: For $N_{train}(t)$ *Individual* agents chosen randomly at time t, increasing permanently their opinion on the "Farm management" topic ($Op^i_{management}$) and their skill regarding the constraint "Technical" ($Skill^i_{technical}$) of a value $\theta_{train}(t)$:

$$\begin{aligned} Op^i_{management}(t+1) \leftarrow & \min(1.0, Op^i_{management}(t) \\ & + \theta_{train}(t)) \end{aligned} \quad (11)$$

$$Skill^i_{technical}(t+1) \leftarrow Skill^i_{technical}(t) + \theta_{train}(t) \quad (12)$$

  The cost of this action is: $N_{train}(t) \times \theta_{train}(t)$
- *Financial support*: For all agents, increases temporary the opinion on the "Economy" topic ($Support^i_{economy}$) of a value $\theta_{aid}(t)$:

$$Support^i_{economy}(t+1) \leftarrow \min(1.0, Support^i_{economy}(t) + \theta_{aid}(t)) \quad (13)$$

  The cost of this action is: $N_{new\_adopters}(t) \times \theta_{aid}(t)$, with $N_{new\_adopters}(t)$ the number of *Individual* agents who have adopted during the policy period t.
- *Environmental awareness*: For $N_{env}(t)$ *Individual* agents chosen randomly, increases permanently the opinion on the "Environment" topic ($Op^i_{environment}$) of a value $\theta_{env}(t)$:

$$Op^i_{environment}(t+1) \leftarrow \min(1.0, Op^i_{environment}(t) + \theta_{env}(t)) \quad (14)$$

  The cost of this action is: $\frac{1}{2} \times N_{env}(t) \times \theta_{env}(t)$. We defined a lower cost for this action because unlike the *training* action which concerns both *Attitude* and *Perceived behavior control*, this one concerns only *Attitude*. Moreover, unlike the *financial support* action which only incurs a cost on the budget if an agent adopts the innovation, here the cost of the action is spent as soon as the action is triggered even if it has no effect on the adoptions.

Table 3 summarizes all the argument values linked to the actions.

At the time of applying an action, the cost of the action is subtracted from the available budget and the action only applied if there is enough budget left. In more detail, for training and

TABLE 3 Parameters that determine the action of the *Institution* agent at time step t.

Action variable	Data type	Description
$N_{train}(t)$	Int [0, 100]	Number of *Individual* agents who will be trained
$\theta_{train}(t)$	Float [0.0, 1.0]	Level of training
$\theta_{aid}(t)$	Float [0.0, 1.0]	Level of financial support
$N_{env}(t)$	Int [0, 100]	Number of *Individual* agents who will be environmentally educated
$\theta_{env}(t)$	Float [0.0, 1.0]	Level of environmental sensibilization

TABLE 4 Parameter values used for the simulations.

State variable	Initial value
N_{ind}	100
N_{social}	5
$P_{interact}$	0.1
Ω_{mean}	0.7
Ω_{std}	0.1
μ	0.1
d	0.5
$Budget_{year}$	10.0
End_{time}	5 years

environmental awareness actions (in this order), the simulation: (i) verifies whether there is enough budget left to apply the action (otherwise the action is not applied at all); and (ii) if it is the case, it applies the action, subtracts its cost from the available budget and continues to process the next type of action if any. Finally, for financial support actions, the aid is offered to new adopters (at the financial level decided by the institution for that period) and its cost subtracted as soon as new individuals adopt until no remaining budget is left or until the period ends.

The model is initialized with the creation of N_{ind} *Individual* agents.

The initialization of an *Individual* agent works as follows: its social network ($social\_network^i$) is filled by N_{social} *Individual* agents chosen at random. Then for each topic k (among Economy, Environment and farm management), the initial value of opinion $Op_k^i(1)$ and weight W_k^i for this topic are chosen randomly between 0.0 and 1.0 (uniform distribution). The weights of the different topics are then normalized so that the sum of the weights is equal to 1.0. The value of TPB weights ($W^i_{attitude}$, W^i_{social} and W^i_{pbc}) are also chosen randomly between 0.0 and 1.0 (uniform distribution) and then normalized. The adoption threshold (Ω^i) is initialized between 0.0 and 1.0 using a truncated Gaussian distribution with a mean of Ω_{mean} and a standard deviation of Ω_{std}. The speed of opinion convergence (μ), the maximal opinion difference accepted for convergence (d) and the probability of interaction ($P_{interact}$) are considered homogeneous for all agents.

Once all these parameters are initialized, the initial value of the Intention is computed using Equation 2.

All (non random) parameters' values as instantiated in experiments are given in Table 4.

5. Policy design using reinforcement learning

This section details our approach that builds on machine learning, and in particular on reinforcement learning, to automatically learn effective policies for innovation diffusion. As initially discussed in the introduction, designing an effective policy for innovation is a very challenging problem. At each decision step, the institution needs to decide if it launches new (parallel) actions on several areas of interest (i.e., training, financial aid and environmental awareness in the particular case study) as well as the extent of each of these actions (i.e., level of increment on the opinion/skill and, for training and environmental awareness actions, the number of individuals reached by the action). Therefore, it is not only which actions to launch but when, in which combination and which parameters' values to choose. This task gets even more complex considering that, to do achieve it, the institution cannot observe any characteristic of the internal state of any individual (e.g., intention, preferences, ...) but only some aggregated statistics at the level of the population (e.g., the number of adopters). The institution can neither target the policy actions to reach specific individuals, but only choose the number of individuals reached.

The remainder of this section describes in detail the approach used: we mathematically formalize the policy design problem faced by the institution as a reinforcement learning problem (Section 5.1) and we describe how to solve it by a deep learning approach (Section 5.2).

5.1. The innovation policy design problem

In what follows we formally define the optimization problem faced by the institution when designing a policy that aims to maximize the number of adopters of an innovation over a finite time horizon, T^{inst} (i.e., the policy will be evaluated for the state reached after T^{inst} institution decision steps). In this problem, the policy actions launched in a given time step t are restricted by the available budget on that time step ($Budget(t)$). We cast this problem in the RL framework in which the institution learns efficient innovation diffusion policies by directly applying action policies on the (simulated) environment. Figure 2 depicts the main steps of interaction occurring between the institution and the environment within a RL iteration.

At every institution decision step $t \in T^{inst}$, the institution receives three observations from the environment (① in Figure 2): the fraction of adopters, the available budget and the number of decision steps remaining before time limit. These observations form the state of the environment perceived by the institution on that time step, i.e., $S(t) =< Budget(t), Adopters(t), Steps(t) >$. Notice that the only indicator that the institution can observe

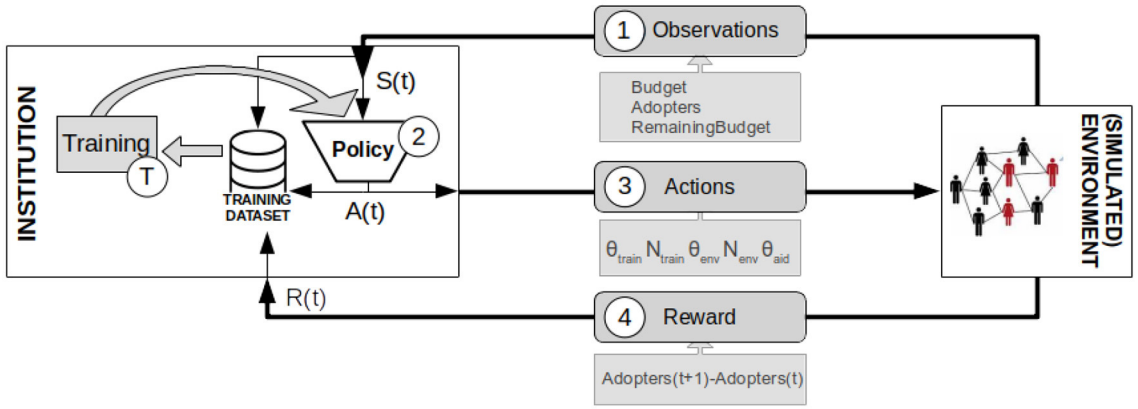


FIGURE 2
Schematic overview of the interactions between the institution (the learner) and the environment during the different RL steps.

from the population is the fraction of adopters. The decision steps remaining before time limit are included to provide time-awareness in the state representation, an approach that has proven good results in time-limited domains [31].

Given the observed state of the environment, the institution applies its policy (② in Figure 2) to decide which actions, $A(t)$, to launch in that state. For this, in this work we consider that the institution needs to define the value of a set of (continuous) parameters for each possible type of action and that multiple types of actions can be launched in parallel[3]. As detailed in Section 4.2, in the particular application studied in this paper, i.e., the adoption of communicating water meters by farmers, there are three types of policy actions: training, financial aid and environmental awareness. In more detail, the institution acts by specifying: (i) the level of each type of action (i.e., the level of training, θ_{train}, the level of the financial aid, θ_{aid} and the level of environmental sensibilisation, θ_{env}); and (ii) the number of individuals to be reached by the action for training (N_{train}) and environmental awareness (N_{env}) actions[4]. Therefore, for this application, the *policy* of the institution will be a stochastic mapping of the following form:

$$\begin{pmatrix} Budget \\ Adopters \\ Steps \end{pmatrix} \rightarrow \begin{pmatrix} \theta_{train} \\ N_{train} \\ \theta_{env} \\ N_{env} \\ \theta_{aid} \end{pmatrix}$$

This stochastic mapping builds on a policy function $\pi(S(t); \Theta)$ which in turn is a parametrized mapping (we use the notation Θ to denote the set of policy parameters) from environmental states to probability distributions over actions. As standard-practice in continuous action spaces, we based our approach on policy-gradient methods in which the institution will learn directly a parameterized policy that can select actions without consulting a value function[5].

Given a policy function, the actions to apply at a given time step t are obtained by directly sampling π on the current state:

$$A(t) \sim \pi(S(t); \Theta) \tag{15}$$

These actions, sent by the institution (③ in Figure 2), are applied to the environment at time step t constrained to the available budget (i.e., an action is only applied if there is enough budget left). In this model we consider that each type of action incurs a cost on the budget which depends on its particular parametrisation and that the institution is aware of the cost of its actions. In Section 4.2, we detailed the costs of each type of action for the particular application studied in this paper.

At time step $t+1$ the environment sends the reward signal[6] observed to the institution (④ in Figure 2). In RL, the learner's sole objective is to maximize the total reward received in the long run. Thus, given that here we are trying to maximize the number of adopters, the reward at a time step t is naturally defined as the increment on the fraction of adopters with respect to the previous step:

$$R(t) = Adopters(t) - Adopters(t-1) \tag{16}$$

Then, for learning purposes, the *Institution* agent stores in its training dataset the information related to this experience, i.e., $< S(t), A(t), R(t+1) >$.

The objective is to find a parametrization of the policy function that maximizes the discounted sum of rewards over time:

$$\max_{\Theta} E\left[\sum_{t=0}^{T^{inst}} \gamma^t \cdot R(t+1) \mid \Theta\right] \tag{17}$$

3 This differs from similar problems studied in the literature where at each time step the agent can only select one action whose parameters are typically decided in a second decision phase.

4 As discussed in Section 4.2, for financial aid the number of individuals reached is not decided ex-ante but determined ex-post since the aid is given iff an individual adopts.

5 A value function may still be used to learn the policy parameters, but is not required for action selection.

6 Reward at time t, $R(t)$ is typically a (stochastic) function of S_{t-1} and A_{t-1}.

where $\gamma \in [0, 1]$ is the discount factor.

At each training step[7] (Ⓣ in Figure 2), the *Institution* agent will input its training dataset to the training module which will output a new set of parameters (i.e., which will be used from this moment to parameterize the current policy in place).

In the next section we describe in detail the design of this training module as well as the particular parametric distributions we propose for the innovation diffusion policy design.

5.2. A deep learning approach to optimize public innovation diffusion policies

This section details the learning approach proposed to optimize the public innovation diffusion policies following the objective detailed in Equation (17). As common in the recent literature when dealing with highly complex partially-observable environments, we based our approach on deep reinforcement learning, where deep stands for an artificial deep neural network (NN) that is used to approximate the policy function. In deep RL, the set of parameters Θ that are adjusted by the learning are the weights of the NN. Also, as typically done in policy parametrization for continuous actions, the policy function is defined as a parametric probability distribution over actions[8] and the outputs of the NN are used to update the parameters of this distribution.

There are several challenges that we need to overcome when designing a neural network policy for this problem. These challenges emerge from the budget constraint that bounds the space of feasible actions depending on the state (i.e., the budget available). The typical approach of dealing with continuous action spaces in deep RL consists in using a normal (Gaussian) policy parametrization in which the NN outputs the mean and the standard deviation of the corresponding normal distribution. If the action is composed of multiple sub-actions, one normal distribution is used to specify each sub-action and the global action distribution is defined as the aggregation of individual distributions. However, notice that if we apply this approach to our problem there is no guarantee that the sampled solution will respect the budget. Still, this does not prevent us from sending these (possibly unfeasible) actions to the environment. As explained in Section 4.2, in this case the actions whose cost exceeds the available budget will not be applied, having the same effect as not launching these actions. However, as highlighted in the literature [20], acting as if there was no constraint in environments with large action spaces typically leads to inefficient learning and poor convergence rate if any. Given this, in this work we opted for an approach which explicitly considers the constraint in the policy, guaranteeing that the actions selected by the policy respect the budget.

Creating a policy that respects the budget constraint is particularly challenging in the innovation policy design problem. In more detail, many constrained problems deal with constraints that apply to individual actions (i.e., define independent bounds on the domain of each sub-action) and, in the continuous action domain, they can be addressed by independently bounding the corresponding distribution of each sub-action [21]. Instead, here the budget is shared among all sub-actions, which in turn can be launched in parallel, so the bounds on an action not only depend on the state but also on the values of other sub-actions. Thus, we can not guarantee that we will respect the constraint by bounding each action individually depending on the state.

Next section describes the particular NN architecture that we propose to overcome the above-mentioned open challenges of using deep RL for innovation diffusion policy design.

5.2.1. A NN architecture to optimize public policies

In the proposed architecture, a classical Gaussian policy parametrisation approach will be used for the probability distributions that define the level with which we apply each type of action at individual level (i.e., θ_{env}, θ_{train} and θ_{aid}). But for the other parameters that characterize the actions, i.e., the number of individuals reached by training and environmental awareness actions, we take a different approach (i.e., those actions will not be directly sampled from probability distributions) that guarantees that the actions selected by the policy respect the budget constraint. In more detail, the NN will output a second set of parameters that define a probability distribution on the allocation of the available budget among the different types of actions. Then, the number of individuals reached by training and environmental awareness actions are unequivocally determined after the realizations of the budget allocation and the level of these actions. Figure 3 depicts this NN architecture along with the different steps of the process that goes from the NN output to the institution policy actions.

5.2.1.1. Normal (Gaussian) distributions for each real-valued θ action

As shown in Figure 3, the NN will output a first set of parameters composed of a mean and a standard deviation for each type of action (e.g., θ_a^{mean}, θ_a^{std} for each $a \in \{train, env, aid\}$). These parameters will be used to define a tanh-squashed normal (Gaussian) distributions over the real-valued action levels. Formally,

$$P(\theta_a|s) = T(\mathcal{N}(\theta_a^{mean}, \theta_a^{std})) \quad \forall a \in \{train, env, aid\} \tag{18}$$

Where T is a tanh-squashed transformation to bound the range of actions ($R \rightarrow [0, 1]$) and $\mathcal{N}$ is the gaussian distribution.

The values of the extent levels of actions are obtained by directly sampling the corresponding distributions on the current state:

$$\theta_{train}(t) \sim P(\theta_{train}|s),\ \theta_{env}(t) \sim P(\theta_{env}|s),\ \theta_{aid}(t) \sim P(\theta_{aid}|s) \tag{19}$$

5.2.1.2. Dirichlet distribution for budget allocation

For modeling the budget allocation among the different types of actions, we propose to use a Dirichlet distribution, a type of distribution typically used in the allocation of (continuous) resources [32] given that the realizations of the distribution satisfy a simplex constraint (i.e., $\sum \mathbf{x} = 1, \mathbf{x} \sim Dir(\cdot)$). Thus, as depicted in Figure 3, the NN will output the concentration parameters of a three-dimensional Dirichlet distribution: one for the training

7 An agent is trained by batches of experiences, the number of experiences per batch being determined by the hyperparameters of the particular learning algorithm used.

8 This parametric distribution over actions can be defined as the aggregation of several parametric distributions on different actions as we do in this work.

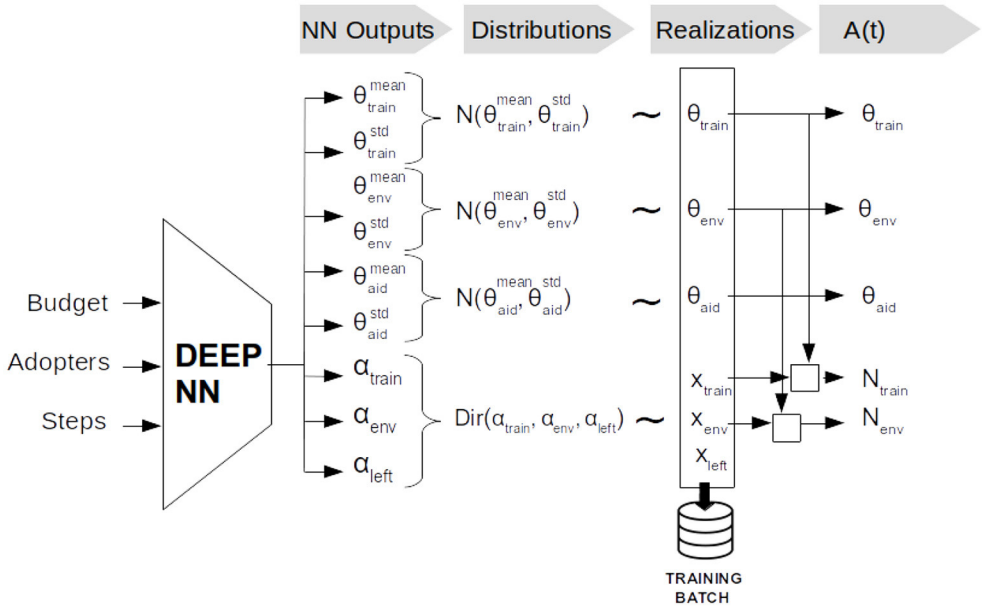


FIGURE 3
Schematic overview of the process from the NN input to the institution policy actions. (1) The NN gets as input the current state of the environment. (2) The outputs of the NN are used to define some parametric distributions over actions in the current state. (3) The distributions are sampled to obtain realizations for the current iteration. (4) Policy actions are defined from the realizations.

dimension (α_{train}), one for the environmental awareness dimension (α_{env}) and one for the budget left for the financial aid dimension and/or to be transferred to the next time step (α_{left}) [9]. Formally,

$$P(x_{train}, x_{env}, x_{left}|s) = Dir(\alpha_{train}, \alpha_{env}, \alpha_{left}) \quad (20)$$

A realization $(x_{train}(t), x_{env}(t), x_{left}(t))$ of the Dirichlet distribution represents a partition of the current budget $Budget(t)$: $(x_{train}(t) \times Budget(t), x_{env}(t) \times Budget(t), x_{left}(t) \times Budget(t))$.

5.2.1.3. Determining N_{env} and N_{train} actions given the budget and θ-actions realizations

Finally, given the current budget partition and the level of extent of each action, the number of individuals reached by training and environmental awareness actions are determined as the maximum number of individuals to which we can apply the selected level for that action while respecting the allocated budget. Formally, $\forall a \in \{train, env\}$

$$N_a(t) = \arg \max_{N \in \{0,\ldots,N_{ind}\}} \left[N \times C_a(\theta_a(t)) \leq x_a(t) \times Budget(t)\right] \quad (21)$$

Where $C_a(\theta_a(t))$ is a function which, given the level of extent to which the action will be applied in the period, returns the cost of applying that action to a single individual. The costs of the actions are specified in Section 4.2, $C_{train}(\theta_{train}(t)) = \theta_{train}(t)$ and $C_{env}(\theta_{env}(t)) = \frac{1}{2} \times \theta_{env}(t)$ for the particular case of study.

Therefore, the Deep NN policy allocates a probability distribution over joint actions to every possible state of the system (year, budget, adopters). This distribution can be sampled whenever we require an action to apply to the system.

6. Results

This section presents the experiments performed to validate the model (Section 6.1) and the framework proposed for policy design using reinforcement learning (Section 6.2) on the adoption of communication water meters in the farmers use case.

The model has been implemented in the open-source platform GAMA [33].[10] The choice of this platform is due to the ease of implementation of models with it, but also, in a perspective of evolution of the model. Indeed, we plan to enrich the model with geographical data to represent real farms and thus be able to calculate in a more advanced way the economic context of the farm and what the communicating water meters can bring them. We also plan to enrich the model by taking up work of Sadou et al. [11] on argumentation to allow a more detailed calculation of the attitude from the knowledge and point of view of the farmer. And GAMA offers integrated tools to support both extensions, namely to integrate geographical data and to explicitly represent arguments in the model [34]. Finally, GAMA also supports communication with external software using message exchange which has facilitated the coupling with the learning module (implemented in Python to take advantage of the existing deep learning libraries). We perform experiments using our implementation of the RL logic and the Proximal Policy Optimization (PPO) algorithm [35] but building on the *Tensorflow* framework[11] for the deep learning part. For the sake of reproducibility, we have made publicly available[12] the source code

9 Note that how the budget left is distributed between the financial aid dimension and budget not spent at time t and thus transferred to time $t+1$ is not controlled by the institution policy.

10 https://gama-platform.org/

11 https://www.tensorflow.org/

12 https://github.com/ptaillandier/policy-design

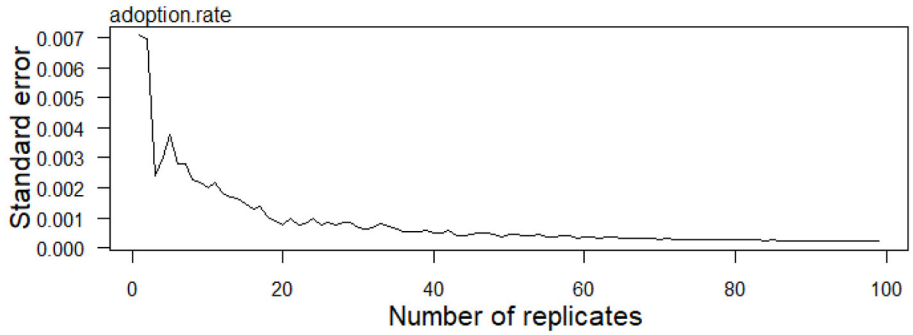


FIGURE 4
Comparison of the standard errors of the adoption rate for the different numbers of replicates.

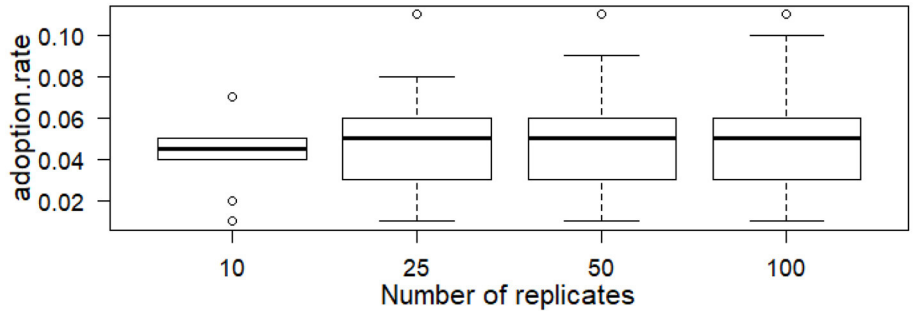


FIGURE 5
Whiskers plots of the adoption rate for the different numbers of replicates. The black lines represent the median values; the boxes represent the interquartile range (IQR), the whiskers represent the minimum/maximum excluding 1.5 IQR outliers. Points are outliers beyond that distance.

of the agent-based model and the reinforcement learning methods used in the experiments.

6.1. Analysis of the model

6.1.1. Stochasticity sensitivity analysis

In a first experiment, we analyze the impact of the stochasticity of the simulations on the results and in particular on the number of adopters. The main objective is to find a threshold value of replications beyond which an increase in the number of replications would not imply a significant marginal decrease of the difference between the results. To do this, we compare the number of adopters at the end of the simulation (i.e., after 5 years) between different numbers of replications of the simulation. We undertake this exploration with the simplest possible scenario, i.e., without any action implemented by the *Institution* agent.

Figure 4 shows the standard error of the adopters rate obtained with different numbers of replicates. Figure 5 shows the impact of the number of replicates on adoption rate: the black lines show the median, the boxes show the second and third quartiles (IQR), the whiskers show the minimum and maximum excluding outliers (simulation results that differ from the median by more than 1.5 times the IQR).

The results suggest that increasing the number of replications beyond 50 does not have a great impact on the aggregate trend of the simulations.

6.1.2. Analysis of the impact of the institution's actions

We propose here to study the impact of the different possible institution's actions individually. In more detail, we calculate, for the three possible types of action defined (financial support, environmental sensibilisation and training), the average values of adoption obtained for 50 replications, depending on the level of these actions. We consider in these experiments that the same action is applied at each institution's action implementation stage (every 6 months) - the action is thus applied 10 times during the simulation period (5 years).

Table 5 presents the results in terms of adoption rate and cost for different levels of financial support. Tables 6, 7 show, respectively the adoption rate and the cost of the training action according to the level of training and proportion of farmers concerned. Finally, Tables 8, 9 show, respectively the adoption rate and the cost of the environmental sensibilisation action according to the level of sensibilisation and proportion of farmers concerned.

A first result is that it is the *training* action that is the most effective in bringing new adopters. This result is not surprising since this action allows the farmer to better understand the interest

TABLE 5 Adoption rate by level of financial support: mean value for the 50 replications and standard deviation in brackets.

θ_{aid}	Adoption rate	Cost
0.0	0.045 (0.02)	4.5 (2)
0.2	0.059 (0.025)	5.9 (2.5)
0.4	0.078 (0.034)	7.8 (3.4)
0.6	0.116 (0.046)	11.6 (4.6)
0.8	0.175 (0.066)	17.5 (6.6)
1.0	0.242 (0.078)	24.2 (7.8)

TABLE 6 Adoption rate by level of training support (θ_{train}) and number of farmers concerned (N_{train}): mean value for the 50 replications and standard deviation in brackets.

N_{train} \ θ_{train}	0.0	0.2	0.4	0.6	0.8	1.0
0	0.045 (0.026)	0.045 (0.026)	0.045 (0.026)	0.045 (0.026)	0.045 (0.026)	0.045 (0.026)
20	0.045 (0.026)	0.184 (0.067)	0.521 (0.198)	0.764 (0.194)	0.801 (0.171)	0.805 (0.179)
40	0.045 (0.026)	0.373 (0.134)	0.736 (0.187)	0.805 (0.176)	0.824 (0.167)	0.85 (0.154)
60	0.045 (0.026)	0.491 (0.181)	0.778 (0.182)	0.809 (0.174)	0.842 (0.162)	0.872 (0.14)
80	0.045 (0.026)	0.558 (0.186)	0.792 (0.179)	0.825 (0.172)	0.85 (0.156)	0.886 (0.133)
100	0.045 (0.026)	0.574 (0.186)	0.804 (0.169)	0.831 (0.163)	0.866 (0.142)	0.893 (0.13)

TABLE 7 Cost by level of training support (θ_{train}) and number of farmers concerned (N_{train}).

N_{train} \ θ_{train}	0.0	0.2	0.4	0.6	0.8	1.0
0	0	0	0	0	0	0
20	0	40	80	120	160	200
40	0	80	160	240	320	400
60	0	120	240	360	480	600
80	0	160	320	480	640	800
100	0	200	400	600	800	1,000

TABLE 8 Adoption rate by level of environmental sensibilisation (θ_{env}) and number of farmers concerned (N_{env}): mean value for the 50 replications and standard deviation in brackets.

N_{env} \ θ_{env}	0.0	0.2	0.4	0.6	0.8	1.0
0	0.045 (0.026)	0.045 (0.026)	0.045 (0.026)	0.045 (0.026)	0.045 (0.026)	0.045 (0.026)
20	0.045 (0.026)	0.062 (0.03)	0.074 (0.032)	0.076 (0.033)	0.077 (0.033)	0.077 (0.033)
40	0.045 (0.026)	0.071 (0.031)	0.078 (0.031)	0.079 (0.032)	0.081 (0.033)	0.084 (0.031)
60	0.045 (0.026)	0.072 (0.031)	0.078 (0.03)	0.079 (0.031)	0.081 (0.032)	0.084 (0.032)
80	0.045 (0.026)	0.073 (0.03)	0.079 (0.03)	0.082 (0.033)	0.084 (0.033)	0.087 (0.032)
100	0.045 (0.026)	0.073 (0.03)	0.08 (0.03)	0.083 (0.033)	0.085 (0.033)	0.088 (0.033)

TABLE 9 Cost by level of environmental sensibilisation (θ_{env}) and number of farmers concerned (N_{env}).

N_{env} \ θ_{env}	0.0	0.2	0.4	0.6	0.8	1.0
0	0	0	0	0	0	0
20	0	20	40	60	80	100
40	0	40	80	120	160	200
60	0	60	120	180	240	300
80	0	80	160	240	320	400
100	0	100	200	300	400	500

of the innovation for managing the farm (and thus, increase her attitude toward adoption) and at the same time to remove the technical obstacles to adoption (increase the perceived behavior control). However, this action is the most expensive, as in this case of application we consider a budget of 10 per year and that for 5 years, the action allows at best to bring the adoption percentage to 18.4% (training of 20% of farmers to improve their level of technicality of 0.2). At the same time, the *financial support* action allows for an average adoption percentage of around 24.2% for a budget of 24.2.

Indeed, the advantage of the *financial support* action is that it is only spent when one farmer adopts the innovation, which greatly limits the cost in the case where few farmers adopt the innovation. Note that this cost could have been much higher if more agents had adopted the innovation.

The *environmental sensibilization* action had a much smaller effect on results with at best a 4% increase in the adopter percentage for a budget of 50.

6.2. Experiment on policy design

In this experiment the *Institution* agent uses (deep) RL to optimize its policy, following the approach and architecture detailed in Section 5. The *Institution* agent uses a two-layer (64 neurons each) feed-forward neural network for the policy approximation. The policy is trained using the Proximal Policy Optimization (PPO) algorithm [35] with the agent learning by batches, each batch containing experiences of 100 complete[13] episodes[14]. Since each complete episode is composed of 10

13 Since our problem is characterized by a fixed finite time horizon there is no need to truncate the episode before its end as happens in other domains.

14 Following RL standard notation, an episode tracks all experiences obtained during the whole time horizon.

TABLE 10 Training hyperparameters.

	Parameter	Value
RL	Discount factor (γ)	0.99
	Advantage function	GAE
	GAE-lambda	0.95
	Training algorithm	PPO
PPO	Sampling horizon	10 (until episode termination)
	Number of learning iterations	30
	Number of episodes per batch	100
	Number of training epochs per update	10
	Number of training minibatches per update/epoch	10
	Clipping ratio	0.2
	Early termination with KL divergence	No
	Mini-batch splitting	Shuffle transitions
	Recomputation of advantages at epoch level	yes
	Separated networks for policy and value	yes
	Optimizer	Adam
	Adam learning rate	0.0003
	Adam epsilon parameter	0.0000007
Policy NN	Activation function for hidden layers	Tanh
	Number of (hidden) fully-connected layers	2
	Fully connected layer dimension	64
	Last layer scaling	0.01
	Initializer (hidden layers)	Orthogonal with gain=1.41
Value NN	Activation function for hidden layers	Tanh
	Number of (hidden) fully-connected layers	2
	Fully connected layer dimension	64
	Last layer scaling	1.0
	Initializer (hidden layers)	Orthogonal with gain=1.41

The *Last layer scaling* parameter rescales the network weights of the last layer after initialization [36].

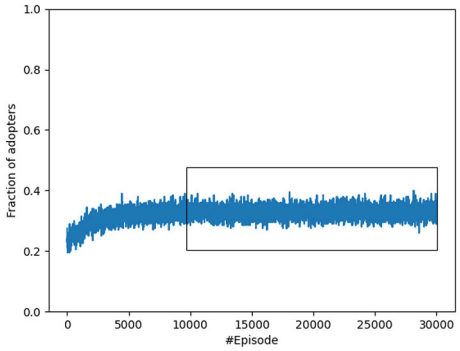


FIGURE 6
Fraction of adopters reached at the end of each simulation episode with the stable training period marked with a box.

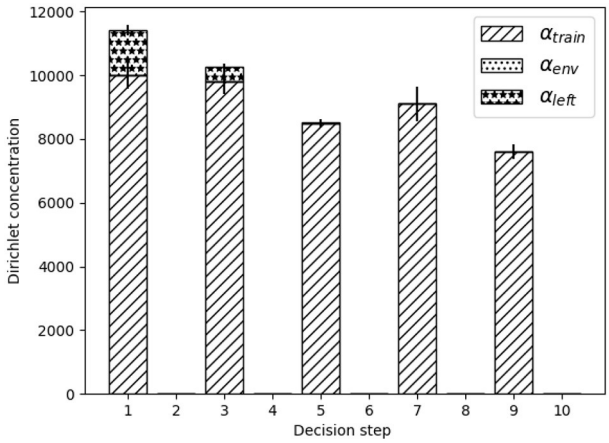


FIGURE 7
Mean values of the Dirichlet concentration parameters, output of the NN to define budget distribution, per institution decision step for the training stable period. Error bars stand for standard deviation.

experiences, this leads to 1000 experiences sampled from the environment in each training iteration using the last policy parameters. In every training iteration, we perform 10 epochs (i.e., the learner will perform 10 passes over the whole batch training dataset). In each epoch, the indices of experiences in the batch are randomly shuffled (i.e., shuffling transitions) and shuffled experiences are partitioned into 10 mini-batches (i.e., each minibatch containing 100 experiences). Table 10 details the training hyperparameters used in our experiments. The hyperparameters were defined following the recommendations issued after the large experimental study carried out for online deep RL methods in Andrychowicz et al. [36]. Regarding the execution time, as it is frequent in deep reinforcement learning the execution time is dominated by the generation of experiences on the simulation module (the simulation of one complete episode[15] takes around 10 s[16]). The time spent to actual training is negligible in comparison (one training iteration takes less than 1 s[16]).

Figure 6 shows the empirical training progress by plotting the adoption rate reached at the end of each episode, resulting from the application of the current policy learned by the institution. The first thing we observe is that the *Institution* agent is able to learn efficient policies (converging to policies with more than 33% of adopters) after 10 learning iterations (recall that with the used parameters the learning occurs every 1000 simulation episodes). For the whole training period the minimum fraction of adopters is 0.09, the maximum is 0.57 and the mean is 0.32 (with standard

15 One complete episode contains 10 decision steps for the institution but also the 365(days)*5(years) simulation steps for individual agents.

16 Experiments where run on an Intel(R) Xeon(R) W-2133 CPU @ 3.60 machine with 128GiB System memory and GPU GeForce RTX 2080.

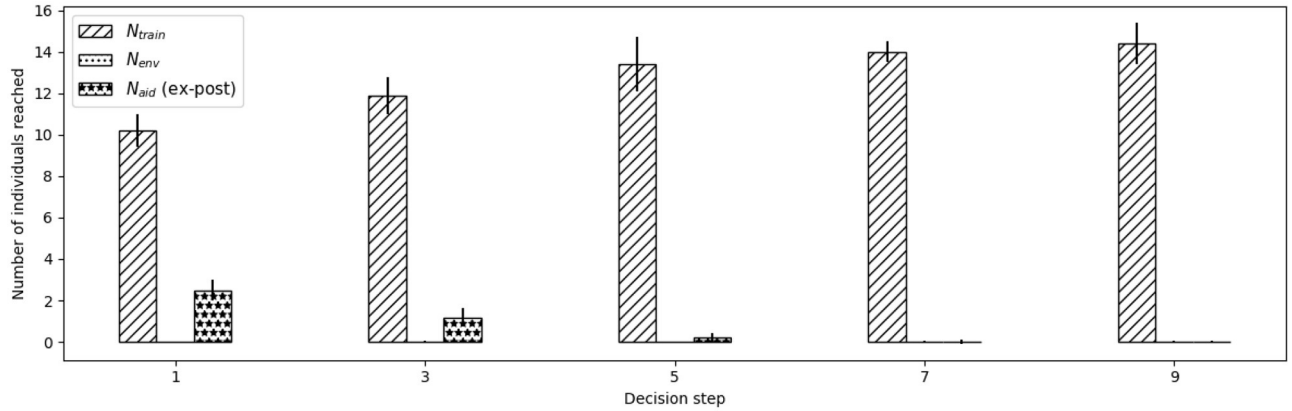


FIGURE 8
Mean values of the N_{train} and N_{env} parameters chosen by the institution and of the observed N_{aid} on odd decision steps during the training stable period. Error bars stand for standard deviation.

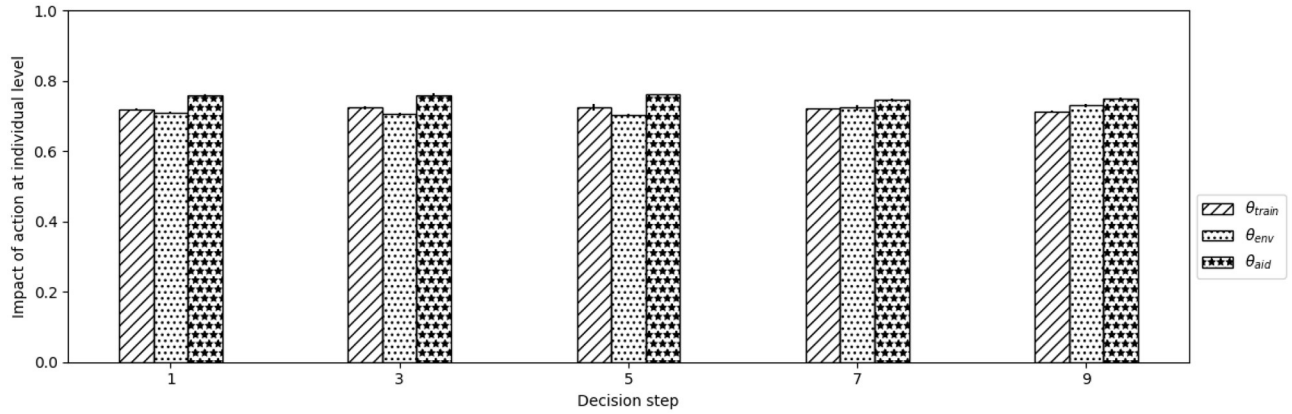


FIGURE 9
Mean values of the level of impact of each type of action (i.e., θ-parameters) chosen by the institution on odd decision steps during the training stable period. Error bars stand for standard deviation.

deviation of 0.0489). We consider that the learning is stabilized after 10,000 training episodes (this with no parallelisation, i.e., all simulations run sequentially, which requires 1–2 days of training[16]) and use the interval 10,000–30,000 to study the structure of the learned policies.

Figure 7 shows, for each institutional decision step, the average of the concentration parameters generated by the NN, i.e., used to define the Dirichlet probability distribution on the budget allocation among the different action types. The first thing we observe in this graph is that concentration values of odd decision steps (corresponding to the beginning of the year when the institution budget gets incremented by 10) are much larger than those of even steps. In fact, this is a consequence of the institution using the budget as soon as it becomes available (i.e., leaving no budget left for future time steps). This is very clear on steps 5, 8 and 9 for which the concentration parameter α_{left} is close to 0. Instead, steps 1 and 3 allocate some budget *via* α_{left} but as we will see later in Figure 8, this budget is not left for the next step but instead entirely spent on the financial support action (i.e., providing the financial aids to new adopters). As a result of this, budget allocation at even steps has no effect on the reward and the NN is unable to reduce the variance of the distribution leading to the observed smaller concentration values.

The second thing we observe in Figure 7 is that the institution learned to spend nearly all budget on the most effective type of action: the training action. As discussed in Section 6.1.2 when evaluating the baselines policies for each type of action, the effectiveness of this type of action is explained as it increases the attitude to adoption at the same time as it increases the perceived behavior control. Despite this dominance of the training dimension, in the first decision steps the institution finds profitable to reserve some budget for ex-post financial aids. Finally, no budget is allocated to environmental sensibilisation actions.

Figures 8, 9 show respectively for the N-parameters and the θ-parameters, the average values of the parameters chosen by the institution at each decision step to parameterize each type of action. For the sake of clarity, we only plot decision steps corresponding to the beginning of a year (i.e., odd time steps) since, as discussed

above, for the rest of the time steps there is no budget left to apply any action to any individual. The first thing we observe in Figure 8 is that, unsurprisingly and as a consequence of the budget allocated, the number of agents reached by environmental awareness actions (N_{env}) is 0 in all decision steps. For financial aids, there is some percentage of individuals that get financial aids on the first steps (i.e., around 2 individuals get the financial aid at step 1 and around a single individual at step 3). As we see in Figure 9 the level of aid (θ_{aid}) proposed to individuals is quite high (values around 0.75). Finally, for training actions, we observe that the level of training does not vary much across the decision steps and it is quite high (values around 0.72). In Figure 8, we observe that this results in around 10-11 individuals trained at each earlier time steps (in which some budget was allocated to financial aids) and in around 14 individuals trained at later time steps (in which the whole budget is allocated to that type of action).

7. Conclusions and future work

This paper proposes an AI framework for the design of innovation diffusion policies. The innovation diffusion policy design problem is a complex sequential decision-making task in which an institution needs to decide which policy actions to launch over time in order to maximize the number of adopters of an innovation after a finite time horizon. The actions available are constrained by the available budget at the decision time. The proposed framework builds on two distinguished components:

- Agent-based simulations, used as a virtual environment in order to conduct a large number of experiments that would be prohibitive on the real environment; and
- (Deep) reinforcement learning, used to automatically identify good-candidate policies in the extremely large search space of possible ones.

In our framework, the agent-based simulations make use of the theory of planned behavior to simulate the behavior of adopters in a credible way, while keeping the computational cost of simulations affordable. Then, a deep reinforcement learning agent interacts with these simulations in order to efficiently explore the exponential space of institutional policies, eventually learning the structure of efficient innovation diffusion policies. The learning represents the policy *via* a neural network architecture that guarantees the respect of the budget constraint by implicitly learning the budget allocation among the different types of actions.

The proposed framework is illustrated in the specific use case of the adoption of communicating water meters by farmers in the Louts region (South-West of France). Empirical results demonstrate the viability and soundness of our approach to identify good-candidate innovation adoption policies for this particular application.

We identify multiple directions that can be pursued as future work. The paper demonstrates for the first time that an AI framework that combines (deep) RL and agent-based simulations is sound and viable for the innovation policy design problem, learning effective policies in the presence of non mutually-exclusive parameterizable actions and budget constraints. However, these results are a first step since any real-support to policymaking will require simulations with ground in real-data as well as widespread consultation with policy-makers. So future research should calibrate the model with real-world data. For the specific use case of communicating water meters we are currently carrying out interviews with farmers in the South-West of France in order to better understand their opinion on these water meters and to be able to move from a random initialization of the agents' attributes to values based on real data. We also hope to get real data on the adoption of these new water meters so that we can also calibrate the model with this information. Second, regarding the NN internal architecture, as future work we plan to test Long Short-Term Memory (LSTM) networks, typically used [5, 6] for encoding the history of past observations in partially-observable environments, in order to analyze the performance impact, instead of the simpler feed forward NN architecture used in this paper. Finally, regarding the objective used in the policy optimization, in this work the budget of the policy maker is represented only as a constraint but not in the objective function. Future work would consider enhancing the reward function in order to minimize the budget as a secondary objective and analyzing the impact of this on the structure of optimal policies.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

Author contributions

The authors have worked together, each in their own specialty, to design and describe the work presented in the paper. MV and PT are the principal designers of this document and are responsible for much of the writing. The agent-based model was mainly developed by PT, LS, and KC. The reinforcement learning approach was developed by MV and RS. BL developed the methods allowing interactions between the simulator and the learning module. Finally, RT and SC, in addition to providing their expertise on the design of models and participated in the writing of the article. All authors contributed to the article and approved the submitted version.

Funding

This work had been funded by INRAE (MathNum department) and by the #Digitag convergence institute (ANR 16-CONV-0004).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Kiesling E, Günther M, Stummer C, Wakolbinger LM. Agent-based simulation of innovation diffusion: a review. *Central Eur J Operat Res*. (2012) 20:183–230. doi: 10.1007/s10100-011-0210-y

2. Bass FM. A new product growth for model consumer durables. *Manag Sci*. (1969) 15:215–27. doi: 10.1287/mnsc.15.5.215

3. Ajzen I. The theory of planned behavior. *Organ Behav Hum Decis Process*. (1991) 50:179–211. doi: 10.1016/0749-5978(91)90020-T

4. Bourceret A, Amblard L, Mathias JD. Governance in social-ecological agent-based models: a review. *Ecol Soc*. (2021) 26:238. doi: 10.5751/ES-12440-260238

5. Zheng S, Trott A, Srinivasa S, Naik N, Gruesbeck M, Parkes DC, et al. The AI economist: improving equality and productivity with AI-driven tax policies. *arXiv preprint* arXiv:200413332 (2020). doi: 10.48550/arXiv.2004.13332

6. Trott A, Srinivasa S, van der Wal D, Haneuse S, Zheng S. Building a foundation for data-driven, interpretable, and robust policy design using the ai economist. *arXiv preprint* arXiv:210802904. (2021) doi: 10.2139/ssrn.3900237

7. Danassis P, Filos-Ratsikas A, Faltings B. Achieving diverse objectives with AI-driven prices in deep reinforcement learning multi-agent markets. *arXiv preprint* arXiv:210606060 (2021). doi: 10.48550/arXiv.2106.06060

8. Liu Y, Halev A, Liu X. Policy learning with constraints in model-free reinforcement learning: a survey. In: Zhou Z, editor. *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI 2021 Virtual Event/Montreal, Canada, 19-27 August 2021*. Montreal, QC: ijcai.org (2021). p. 4508–15.

9. Deffuant G, Huet S, Amblard F. An individual-based model of innovation diffusion mixing social value and individual benefit. *Am J Sociol*. (2005) 110:1041–69. doi: 10.1086/430220

10. Deffuant G, Neau D, Amblard F, Weisbuch G. Mixing beliefs among interacting agents. *Adv Complex Syst*. (2000) 3:87–98. doi: 10.1142/S0219525900000078

11. Sadau L, Couture A, Thomopoulos R, Taillandier P. Better representing the diffusion of innovation through the theory of planned behavior and formal argumentation. In: *Advances in Social Simulation: Proceedings of the 16th Social Simulation Conference*. Springer (2022). p. 423–35.

12. Rogers EM. *Diffusion of Innovations*. 5th ed. New York, NY: Free Press (1962).

13. Zhang H, Vorobeychik Y. Empirically grounded agent-based models of innovation diffusion: A critical review. *Artif Intell Rev*. (2019) 52:707–41.

14. Bourceret A, Amblard L, Mathias JD. Adapting the governance of social-ecological systems to behavioural dynamics: an agent-based model for water quality management using the theory of planned behaviour. *Ecol Econ*. (2022) 194:107338. doi: 10.1016/j.ecolecon.2021.107338

15. Beedell J, Rehman T. Using social-psychology models to understand farmers' conservation behaviour. *J Rural Stud*. (2000) 16:117–27. doi: 10.1016/S0743-0167(99)00043-1

16. Masson W, Ranchod P, Konidaris GD. Reinforcement learning with parameterized actions. In: Schuurmans D, Wellman MP, editors. *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016*. Phoenix, AZ:AAAI Press (2016). p. 1934–40.

17. Hausknecht MJ, Stone P. Deep reinforcement learning in parameterized action space. In: Bengio Y, LeCun Y, editors. *4th International Conference on Learning Representations, ICLR 2016. San Juan, Puerto Rico, May 2-4, 2016 Conference Track Proceedings*. San Juan (2016).

18. He J, Ostendorf M, He X, Chen J, Gao J, Li L, et al. Deep reinforcement learning with a combinatorial action space for predicting popular reddit threads. In: Su J, Carreras X, Duh K, editors. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016 Austin, Texas, USA, November 1-4, 2016*. Austin, TX: The Association for Computational Linguistics (2016). p. 1838–48.

19. Delarue A, Anderson R, Tjandraatmadja C. Reinforcement learning with combinatorial actions: an application to vehicle routing. In: Larochelle H, Ranzato M, Hadsell R, Balcan M, Lin H, editors. *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020 NeurIPS 2020 December 6-12, 2020*. (2020). Available online at: https://proceedings.neurips.cc/paper/2020/hash/06a9d51e04213572ef0720dd27a84792-Abstract.html

20. Huang S, Ontañón S. A closer look at invalid action masking in policy gradient algorithms. In: Barták R, Keshtkar F, Franklin M, editors. *Proceedings of the Thirty-Fifth International Florida Artificial Intelligence Research Society Conference, FLAIRS 2022 Hutchinson Island, Jensen Beach, Florida, USA, May 15-18, 2022*. Jensen Beach, FL (2022).

21. Chou P-W, Maturana D, Scherer SA. Improving stochastic policy gradients in continuous control with deep reinforcement learning using the beta distribution. In: Precup D, The YW, editors. *Proceedings of the 34th International Conference on Machine Learning, Vol.70*. Sydney, NSW: PMLR (2017). p. 834–43. Available online at: http://proceedings.mlr.press/v70/chou17a.html

22. Dalal G, Dvijotham K, Vecerík M, Hester T, Paduraru C, Tassa Y. Safe exploration in continuous action spaces. *CoRR*. (2018) abs/1801.08757. doi: 10.48550/arXiv.1801.08757

23. Bhatia A, Varakantham P, Kumar A. Resource constrained deep reinforcement learning. In: Benton J, Lipovetzky N, Onaindia E, Smith DE, Srivastava S, editors. *Proceedings of the Twenty-Ninth International Conference on Automated Planning and Scheduling, ICAPS 2018*. Berkeley, CA: AAAI Press (2019). p. 610–20.

24. Chow Y, Nachum O, Duenez-Guzman EA, Ghavamzadeh M. A lyapunov-based approach to safe reinforcement learning. In: Bengio S, Wallach HM, Larochelle H, Grauman, Cesa-Bianchi N, Garnett R, editors. *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018*. Montreal, QC (2018), p. 8103–12. Available online at: https://proceedings.neurips.cc/paper/2018/hash/4fe5149039b52765bde64beb9f674940-Abstract.html

25. Liu Y, Ding J, Liu X. IPO: Interior-point policy optimization under constraints. *Proc AAAI Conf Artif Intell*. (2020) 34:4940–7. Available online at: https://ojs.aaai.org/index.php/AAAI/article/view/5932

26. Yang T, Rosca J, Narasimhan K, Ramadge PJ. Projection-based constrained policy optimization. In: *8th International Conference on Learning Representations, ICLR 2020 Addis Ababa, Ethiopia, April 26-30, 2020*. Ababa: OpenReview.net (2020).

27. Grimm V, Railsback SF, Vincenot CE, Berger U, Gallagher C, DeAngelis DL, et al. The ODD protocol for describing agent-based and other simulation models: a second update to improve clarity, replication, and structural realism. *J Artif Soc Soc Simulat*. (2020) 23:7. doi: 10.18564/jasss.4259

28. Stoneman P, Diederen P. Technology diffusion and public policy. *Econ J*. (1994) 104:918–30. doi: 10.2307/2234987

29. Collard A-L, Garin P, Montginoul M. Un compteur guillemotleft intelligent guillemotright pour mesurer les usages de l'eau: l'entree en scene d'une nouvelle connaissance. *Developpement durable et territoires, Economie, geographie, politique, droit, sociologie*. (2019) 10.

30. Sadou L, Couture S, Thomopoulos R, Taillandier P. Simuler la diffusion d'une innovation agricole à l'aide de modèles à base d'agents et de l'argumentation formelle. *Revue Ouverte d'Intelligence Artificielle*. (2021) 2:65–93. doi: 10.5802/roia.10

31. Pardo F, Tavakoli A, Levdik V, Kormushev P. Time limits in reinforcement learning. In: *Proceedings of the 35th International Conference on Machine Learning, Vol. 80*. Stockholm: PMLR (2018). p. 4042–51. Available online at: http://proceedings.mlr.press/v80/pardo18a.html

32. Tian Y, Han M, Kulkarni C, Fink O. A prescriptive Dirichlet power allocation policy with deep reinforcement learning. *Reliabil Eng Syst Safety*. (2022) 224:108529. doi: 10.1016/j.ress.2022.108529

33. Taillandier P, Gaudou B, Grignard A, Huynh QN, Marilleau N, Caillou P, et al. Building, composing and experimenting complex spatial models with the GAMA platform. *GeoInformatica*. (2019) 23:299–322. doi: 10.1007/s10707-018-00339-6

34. Tailandier P, Salliou N, Thomopoulos R. Introducing the argumentation framework within agent-based models to better simulate agents' cognition in opinion dynamics: application to vegetarian diet diffusion. *J. Artif. Soc. Soc. Simul*. (2021) 24:1–6. doi: 10.18564/jasss.4531

35. Schulman J, Wolski F, Dhariwal P, Radford A, Klimov O. Proximal policy optimization algorithms. *arXiv preprint* arXiv:170706347 (2017). doi: 10.48550/arXiv.1707.06347

36. Andrychwicz M, Raichuk A, Stanczyk P, Orsin M, Girgin S, Marinier R, et al. "What matters for on-policy deep actor-critic methods? A large-scale study," in *9th International Conference on Learning Representation*. OpenReview.net (2021). Available online at: https://openreview.net/forum?id=nIAxjsniDzg

Frontiers in
Applied Mathematics and Statistics

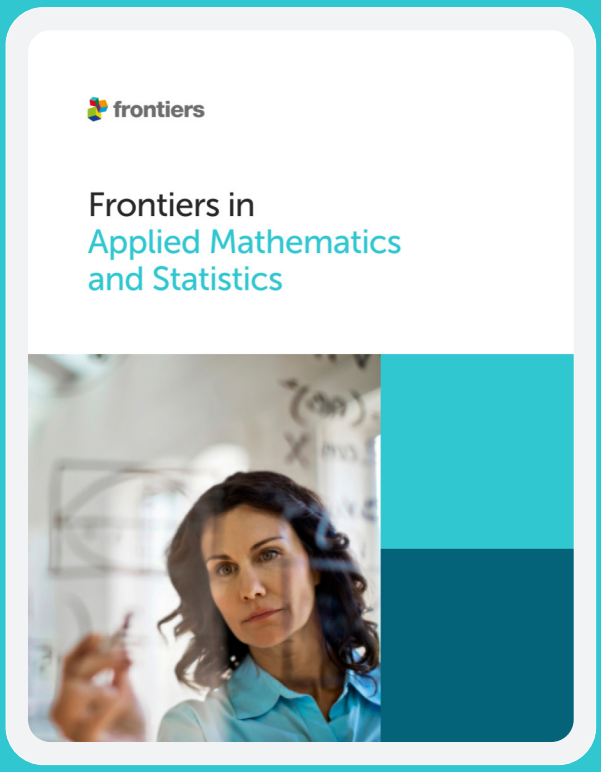
Investigates both applied and applicable mathematics and statistical techniques

Explores how the application of mathematics and statistics can drive scientific developments across data science, engineering, finance, physics, biology, ecology, business, medicine, and beyond

---

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact