Modeling and Simulation in Science,
Engineering and Technology

Nicola Bellomo
Mark A. J. Chaplain
Editors

Predicting Pandemics in a Globally Connected World, Volume 1

Toward a Multiscale, Multidisciplinary Framework through Modeling and Simulation

Birkhäuser

Modeling and Simulation in Science, Engineering and Technology

Series Editors

Editorial Board Members

Nicola Bellomo • Mark A. J. Chaplain
Editors

Predicting Pandemics in a Globally Connected World, Volume 1

Toward a Multiscale, Multidisciplinary Framework through Modeling and Simulation

Birkhäuser

Editors
Nicola Bellomo
Department of Mathematical Sciences
Politecnico di Torino
Torino, Italy

Mark A. J. Chaplain
School of Mathematics and Statistics
University of St. Andrews
St. Andrews, UK

ISSN 2164-3679 ISSN 2164-3725 (electronic)
Modeling and Simulation in Science, Engineering and Technology
ISBN 978-3-030-96561-7 ISBN 978-3-030-96562-4 (eBook)
https://doi.org/10.1007/978-3-030-96562-4

This book is published under the imprint Birkhäuser, www.birkhauser-science.com by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

The onset of the *SARS–CoV–2* virus, responsible for the initial *COVID-19* outbreak and the subsequent pandemic, has brought to almost all countries across the globe enormous problems affecting health, safety, economics and, more generally, all expressions of collective behaviour in our societies. The need to tackle this crisis situation has promoted a huge literature in the field, and some advanced contributions have traced new challenging research lines that have effectively contributed to a deeper understanding of the complex system under consideration.

This intense activity has motivated the editing of this book, which presents 10 chapters devoted to different aspects of the interaction of mathematics with biology, virology and immunology. Research activity has also focused on the impact of the pandemic on economics, social sciences and welfare policy. The awareness of the role that science can have to contribute to fighting the pandemic has grown rapidly not only among scientists but also in the society and governments.

In particular, the mathematical sciences (specifically modelling and computational simulation) can contribute to understanding the scenarios of outcomes that follow the onset of a pandemic depending on the actions that are developed at a local, national and international level. Therefore, mathematics can contribute to the decision making of crisis managers in charge of planning actions to weaken the damage of the pandemic, and policymakers who are tackling all related social and economic problems. These reasonings cannot be confined to mathematics alone since, according to our opinion, one requires a coordinated, and hopefully, synergetic interaction of all the so-called hard sciences. It is to be hoped that from fruitful and virtuous interactions one might end up with devices of artificial intelligence where models and computing interact with data in order to support crisis managers and further developments of research activity.

The contents of the various contributions to this edited book are described in the first "editorial" chapter which refers to key research objectives. Therefore, we do not repeat the contents but remark on the key features of the contributions. In more detail, five key topics are brought to the readers' attention:

1. *An interdisciplinary vision, where the impact of the pandemic on social and economic problems is deeply studied*
2. *A search for new mathematical tools towards modelling and simulations developed within a multiscale vision*
3. *The modelling of spatial dynamics, considering both networks and spread in continuous environments*
4. *A study and interpretation of real data related to well-defined regional and state environments*
5. *Analytic and computational methods generated to the application of models which account also for control problems*

According to these reasonings, we trust that this book can guide future research activity in the field.

Granada, Spain — Nicola Bellomo
St. Andrews, Scotland, UK — Mark A. J. Chaplain
November 2021

Contents

Modelling, Simulations, and Social Impact of Evolutionary Virus Pandemics

Nicola Bellomo, Diletta Burini, Mark A. J. Chaplain, and Nisrine Outada

1 Aims and Plan of the Chapter

The onset of the *SARS–CoV–2* virus responsible for the initial *COVID-19* outbreak and the subsequent pandemic has brought to almost all countries across the globe huge problems affecting health, safety/security, economics, and practically all expressions of collective behaviour in our societies heavily oppressed throughout the world. Indeed, the pandemic has generated severe problems affecting health, but also the economies in our societies have been affected thus requiring a new vision of social organization and a greater attention to welfare problems.

On the other hand, the pandemic has contributed to improving individual collaboration, against selfishness, and collective trust in the role of science including the mathematical sciences, specifically, mathematical modelling and simulations.

N. Bellomo (✉)
University of Granada, Granada, Spain

IMATI CNR, Pavia, Italy
e-mail: bellomo@imati.cnr.it

D. Burini
University of Perugia, Perugia, Italy
e-mail: dilettaburini@alice.it

M. A. J. Chaplain
Mathematical Institute, School of Mathematics and Statistics, University of St Andrews, St Andrews, Scotland, UK
e-mail: majc@st-andrews.ac.uk

N. Outada
Faculty of Sciences Semlalia LMDP, UMMISCO (IRD-SU), Cadi Ayyad University, Marrakech, Morocco
e-mail: nisrine.outada@uca.ac.ma

N. Bellomo, M. A. J. Chaplain (eds.), *Predicting Pandemics in a Globally Connected World, Volume 1*, Modeling and Simulation in Science, Engineering and Technology, https://doi.org/10.1007/978-3-030-96562-4_1

These can contribute to understanding the scenarios of the outcomes that follow the onset of a pandemic depending on the actions that are developed at a local, national, and international level. Therefore, mathematics can contribute to the decision making of crisis managers in charge of planning actions to weaken the damage of the pandemic, and policy makers who are tackling all related social and economic problems. However, these reasonings cannot be confined to mathematics alone since, according to our opinion, one requires a coordinated and, hopefully, synergetic interaction of all the so-called hard sciences. It is to be hoped that from fruitful and virtuous interactions one might end up with devices of artificial intelligence where models and computing interact with data in order to support crisis managers and further developments of research activity.

The overall situation has promoted an enormous, one may even say excessive, literature in the field, where specific contributions, selected among several ones, have traced new challenging research lines which have contributed to a deeper understanding of the complex system under consideration. This general state of the art has motivated the editing of this book which presents 10 chapters devoted to different aspects of the interaction not only between mathematics and biology, virology, immunology but also economics, social sciences, and welfare policy. This interdisciplinary vision was the guiding hint of [8] which has been, subsequently, followed by various authors including those of this edited book.

This editorial chapter aims not only at presenting the contents of the following chapters but also at identifying some research perspectives that will engage applied mathematicians within the framework of the aforementioned interdisciplinary vision. The search for scientific perspectives can be related to a selection of hints that can guide future research activity in the field. In more details, we identify five key topics—KT—selected according to the authors' bias and research knowledge that might guide future mathematical contributions.

KT1 The modelling approach should go far beyond deterministic population dynamics, as individual reactions to the infection and pandemic events are heterogeneously distributed throughout the population. Different types of heterogeneity appear in addition to that corresponding to biology. For instance physical, related to age, or social including the level of education which leads to a greater or lower level of ability to reach an advanced level of awareness concerning the risk of contagion [22, 23, 25].

KT2 Modelling should be developed according to a multiscale vision. Indeed, the pathology of individuals depends on the within-host dynamics at a small scale inside the lung. The dynamics concern also the competition between virus particles and the immune system. Interactions between individuals, either infected or susceptible, occur at a large scale, which, however, is related to the small scale as the infection rate depends on the viral charge. This topic is studied in a parallel field of biology—see the multiscale approach to modelling pattern formation in cancer phenomena [26], and the modelling between cancer and immune cells [12].

KT3 Spatial dynamics play an important feature and have to be included in the overall approach. These should account not only for the dynamics of interactions in crowds and, more generally, the aggregation of individuals but also for interactions induced by transportation networks since we live in a globally connected world. Pursuing this objective leads to new concepts of networks that require to be considered in the complex organization of human societies referring to their transportation systems and in some way related to social interactions including the organization of the broad variety of social activities.

KT4 Models should consider the aspects of post-Darwinist dynamics with mutations and selection. Viruses can mutate by errors in the reproduction dynamics. Most of these mutations generate individual entities which are suppressed since they are less fit with regard to the environment, but some generate fitter variants with higher reproductive ability. The modelling should consider the overall complexity of the in-host competition as well as therapeutical actions that can mitigate the aggressiveness of the virus and/or the activation of the immune system.

KT5 As observed in [10], the modelling of the in-host competition should consider that the immune system is constituted by several interacting populations each of which can progress and proliferate. Each population has a well-defined role and some of them are messengers carrying different types of information [18]. The literature in the field is growing rapidly, see [14, 27], as a natural development of classical theories [15], while the immune system evolves by learning about the presence of the "unwelcome guest".

The list given above does not claim to be complete. However, it already identifies a number of highly challenging perspectives. Some of them have been already introduced in the pioneering article [8] which is considered in some of the chapters of this special issue. The description of the contents of this book is delivered in Sect. 2, while Sect. 3 focuses on a forward look to research perspectives.

2 On the Contents of the Edited Book

This section presents the contents of the chapters which follow this editorial chapter in the edited book. The interested reader can identify, already by a concise description of the contents, how each chapter tackles, at least in part, the key topics (hints) indicated in Sect. 1. Here we simply mention that all KTs are somehow interconnected among themselves as we will critically overview in the next section looking ahead to research perspectives.

Chapter 2 presents two different approaches to modelling the spread of COVID-19 focusing on the current epidemiological situation of vaccination roll-out [1]. The first approach consists in a stochastic SHARUCD modelling framework which developed with a modelling task force created to support public health managers during the COVID-19 crisis. Validation is treated on the basis of empirical data

available for the Basque Country, Spain. The second approach is based on the kinetic theory of active particles (KTAP) model for the spread of a disease.

Chapter 3 deals with the mathematical modelling of epidemic phenomena by means of kinetic equations [4]. The following topics are treated: (i) the role of social characteristics of the interacting individuals, such as the age of individuals, the number of social contacts, and their economic wealth, played in the spread of an epidemic; (ii) optimal control formulation, based on the policy makers' perception of the progress of the epidemic, to contain the epidemics; (iii) multiscale transport settings to model spatial dynamics on both one-dimensional networks and realistic two-dimensional geographic settings.

Chapter 4 provides, as in [5, 8], a sharp analysis of the complexity and multiscale features of COVID-19 pandemics. This study specifically focuses on the spread of the virus in Hawaii, with natural geography barriers between the major cities, and on the spread in New Jersey, where population centres are tightly connected within the state, and the neighbouring major cities, New York and Philadelphia. Then, out of this analysis, the authors study the advantages and limitations of agent-based and compartmental transmission models which are referred to the specific geographic and socio-demographic characteristics of the aforementioned regions. This sharp analysis leads to quantify the predictive ability of models.

Chapter 5 During the COVID-19 pandemic, conflicting opinions on physical distancing swept across social media, affecting both human behaviour and the spread of COVID-19. Inspired by such phenomena, we construct a two-layer multiplex network for the coupled spread of a disease and conflicting opinions. We model each process as a contagion. On one layer, we consider the concurrent evolution of two opinions—pro-physical-distancing and anti-physical-distancing—that compete with each other and have mutual immunity to each other. The disease evolves on the other layer, and individuals are less likely (respectively, more likely) to become infected when they adopt the pro-physical-distancing (respectively, anti-physical-distancing) opinion.

Chapter 6 The multiscale nature of virus dynamics is exemplified by a detailed discussion of the link between macroscopic compartmental models and microscopic individual based models [19]. Special attention is paid to uncertainty quantification, calibrating models by real data, and to interactions between humans which generate virus diffusion. These interactions concern a much larger scale and enhance complexity and are source of uncertainty. Then, the authors move to the information that can be achieved by macroscopic compartmental models focusing on different scales where the dynamics develop depending also on the geographic features of the areas, where the epidemic develops.

Chapter 7 introduces an agent-based model of a workplace where the mood (i.e., the motivation to work) improves when the agent is socializing and deteriorates when it is working [20]. The modelling considers new concepts of social distancing which are different from the physical distancing measures and investigates their effects on peoples' mental well-being and the subsequent influence on work productivity. Heterogeneity is considered by modelling that low motivated agents seek company and high motivated agents focus on work. The study shows how

decreasing the frequency or the quality of the social interactions lead the agents to enter a burnout regime, where their motivations diverge to negative infinity, with the consequent vanishing of productivity.

Chapter 8 develops a study on the interplay between transmission dynamics and quarantine intervention of the COVID-19 outbreak in a cluster community [21]. A novel Transmission-Quarantine epidemiological model by non-linear ordinary differential equations system is proposed. Real epidemiological data from the Cruise ship "Diamond Princess" are used to design a Transmission-Quarantine work-flow which determines the optimal case-specific parameters and validate the proposed model by comparing the simulated curve with the real data. Firstly, a general SEIR-type epidemic model is applied to study the transmission dynamics of COVID-19 without quarantine intervention, and subsequently the more general case is studied in view of actions to mitigate the effects of the epidemics.

Chapter 9 focuses on the modelling and simulation of an infectious disease spreading in a medium size population occupying a confined environment, such as an airport terminal [24]. The authors develop a kinetic type model for active particle methods in crowd dynamics, see [7]. The resulting kinetic model features a variable that denotes the level of exposure to people spreading the disease, a parameter describing the contagion interaction strength, and a kernel function that is a decreasing function of the distance between a person and a spreading individual. Computational tests and additional analysis on the modelling of the contagion dynamics complete the contents of the chapter.

Chapter 10 studies modelling and computational aspects of a model which accounts for time-dependent susceptible-exposed-infectious-recovered-died reaction-diffusion system for the virus pandemics [28]. Macroscopic models are derived from the underlining description at the microscopic scale by a multiscale decomposition which leads to an equivalent formulation of the kinetic theory model which couples the microscopic equations with the macroscopic equations. An asymptotic preserving numerical scheme to solve the equivalent formulation is developed and validated by various numerical tests.

3 Reasonings on Research Perspectives

The chapters presented in Sect. 2 have already delivered interesting contributions to the key problems proposed in Sect. 1. These research results focus on multiscale problems, calibration of models by real data analysis, control problems related actions to contain crisis situations, micro-macro derivation, and various other topics. It is interesting to observe that all authors have shown a deep awareness of the complexity features of the biological-social system under consideration and, consequently, of the need to develop new mathematical tools and conceptual approaches rather than blindly using what is already available in the literature.

Accordingly, we can state that the scientific research is actively operating on the challenging topics under consideration and that advanced results have been

achieved. Indeed, these chapters show that new frontiers of applied mathematics have been opened by the complex interaction between mathematical sciences and the time-space dynamics of pandemics driven by a mutating virus. It would be important, also for the benefit of our society, that the scientific effort stimulated by the pandemics will continue into the future. In fact, a spectrum of new mutations or different types of virus could, unfortunately, still appear.

Looking ahead to research perspectives, we can enforce, once more, what is stated in [8] as well as in the editorial paper [9]. Specifically, we refer to the concepts stated in the following:

> The modelling approach should go far beyond a deterministic description of the dynamics. A modelling framework ought to be developed within a multiscale vision, as the dynamics of individuals depend on the dynamics at smaller scales inside each individual by the competition between virus particles and the immune system by individual reactions to the infection and pandemic events which are heterogeneously distributed throughout the population.

A key target of the modelling approach is the study of the *within-host dynamics* which develops, mainly in the lungs, after contagion. It mainly consists in the competition between the immune system and a proliferative virus.

The study of these specific dynamics can provide important information on the evolution of the pathology accounting for the heterogeneity of the immune defence as well as of possible therapeutical actions. Indeed, we feel comfortable to state that this challenging topic deserves being considered as a key perspective of future research activity.

Bearing all of the above in the mind, five key research perspectives **KRPs**, somehow related the five key topics (KTs) are selected, among several possible ones, and brought to the attention of the interested reader. Before providing some hints towards the **KRPs** that we bring to the attention of the interested reader, it may be useful looking at the flow charts in Figs. 1 and 2 which are somehow inspired to [11].

Figure 1 shows the different paths leading to infection which can be mitigated by vaccination programs. The figure shows that infections can be contracted either by different types of localized clusters, such as family clusters and in crowds, as well as by short or long range transportation of individual through networks.

Figure 2 shows the different paths by which infected individuals might encounter three possible outcomes, i.e., direct recovered healthy state, direct death, or hospitalization. Once hospitalized, individuals can again move to healthy state, death, or intensive care hospitalization. The output of this additional passage will end up with one of the two aforementioned extreme outputs.

- **KRP1:** The modelling of the in-host competition refers to the dynamics of Fig. 1. Progress with respect to the existing literature should consider that the immune system is constituted by several interacting populations each of which can progress and proliferate. Each population has a well-defined role and some of them are messengers of different types of information [18]. The literature in the field is growing rapidly, see [14, 27], as a natural development of classical

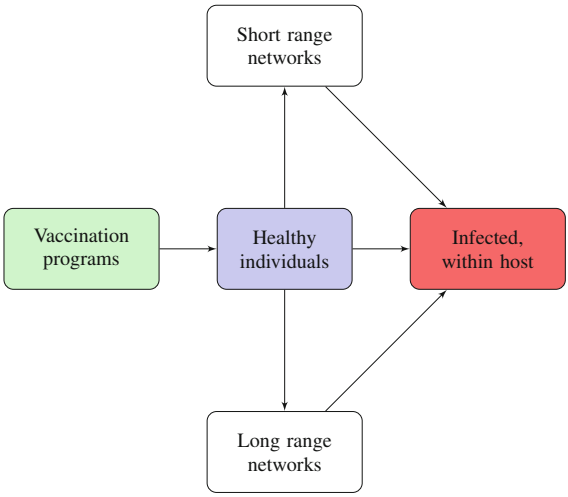


Fig. 1 Flow chart on the path from contagion to infection mitigated by vaccination programs

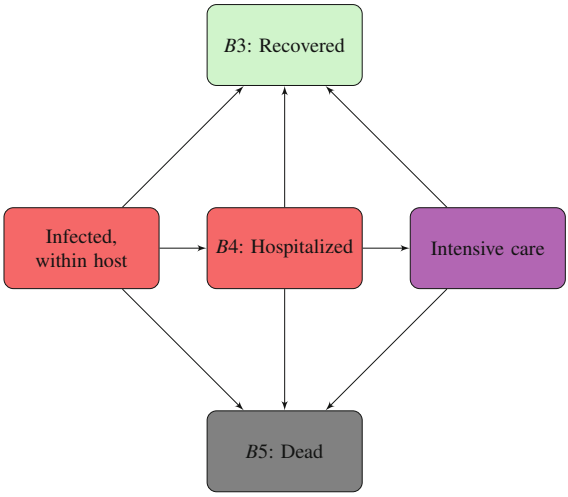


Fig. 2 Flow chart on the path from infection medical care with different intensity of care

theories [15], while the immune system evolves by learning about the presence of the *not-welcome guest*. The first step towards this topic is proposed in [11].

- **KRP2:** A key challenge consists in transferring the biological knowledge into a mathematical framework. A further set of dynamics to be considered is the pseudo-Darwinian mutations of the virus, from the onset to aggressive variants [16] by dynamics somehow related to the lower molecular scale. The in-host dynamics is evolutionary in nature, since the virus evolves by pseudo-Darwinian dynamics. Multiscale topics are treated in Chaps. 2 and 6, see [1, 19], while Chap. 10 shows how models at the macroscopic scale can be derived from the underlying description at the microscopic scale.

- **KRP3:** The study and modelling of medical therapies are definitely worth developing with the aim of showing how these actions can mitigate the in-host proliferation and spatial propagation of the virus. Vaccination strategies [13] should be included in the modelling approach to understand how this action can reduce the spread through a given territory. This action on its own is not sufficient, but it should be coupled with other actions—for instance (at least partial) confinement. In fact the virus is highly resilient and even is minimized, it has a strong ability to survive and grow again as soon as actions to counter it weaken.
- **KRP4:** Contagion problems, shown in Fig. 1, are due to different types of interactions. Spatial dynamics can develop for people moving in open or closed venues [23]. This topic is studied in Chap. 9, see [24], or in family clusters as studied in Chap. 8, see [21]. In addition, a contagion may be caused by local (clusters individuals in crowds), short range (daily movement due to employments in different places of towns and regions), and long distance transportation. Recent literature has started to study this topic intensively as shown in [4].
- **KRPs5:** A further key topic to be studied is the complex interaction between pandemics and society, specifically to economics, social dynamics, and individual behaviours [2]. This topic is treated in Chap. 7, see [20], by development of agent methods. It is worth mentioning that this chapter can conceptually contribute to developments of active particles methods to modelling social dynamics [3] up to recent studies on the complex interactions between pandemics and evolutionary economics [17].

All of the above reasonings lead to the concept that an interdisciplinary vision is necessary through mutually enriching and beneficial interactions with scientists in other fields including virology, epidemiology, immunology—namely biology in general. Not only this but also social sciences and economics as the COVID-19 pandemic has highlighted the fragility of our society [6] and of the whole planet. Research activity in this area has a real role to play to save the planet and future generations.

References

1. M. Aguiar, V. Anam, N. Cusimano, D. Knopoff, N. Stollenwerk, Understanding COVID-19 epidemics: a multi-scale modeling approach, in *Predicting Pandemics in a Globally Connected World, Vol. 1*, *Modeling and Simulation in Science, Engineering, and Technology*, Chap. 2 (Birkhäuser, New York, 2022)
2. M. Aguiar, G. Dosi, D.A. Knopoff, M.A. Virgillito, A multiscale network-based model of contagion dynamics: heterogeneity, spatial distancing and vaccination. Math. Models Methods Appl. Sci. **31**(12), 2425–2454 (2021). Open Access
3. G. Ajmone Marsan, N. Bellomo, L. Gibelli, Stochastic evolutionary differential games toward a systems theory of behavioral social dynamics. Math. Models Methods Appl. Sci. **26**, 1051–1093 (2016)

4. G. Albi, G. Bertaglia, W. Boscheri, G. Dimarco, L. Pareschi, G. Toscani, M. Zanella, Kinetic modelling of epidemic dynamics: social contacts, control with uncertain data, and multiscale spatial dynamics, in *Predicting Pandemics in a Globally Connected World, Vol. 1*, *Modeling and Simulation in Science, Engineering, and Technology*, Chap. 3 (Birkhäuser, New York, 2022)
5. S. Allred, M. Chyba, J.M. Hyman, Y. Mileyko, B. Piccoli, COVID-19 pandemic evolution in Hawai'i and New Jersey: a lesson on infection transmissibility and the role of human behavior, in *Predicting Pandemics in a Globally Connected World, Vol. 1*, *Modeling and Simulation in Science, Engineering, and Technology*, Chap. 4 (Birkhäuser, New York, 2022)
6. B. Avishai, The pandemic isn't a black swan but a portent of a more fragile global system. The New Yorker, 21 April 2020. https://www.newyorker.com/news/daily-comment/the-pandemic-isnt-a-black-swan-but-a-portent-of-a-more-fragile-global-system
7. B. Aylaj, N. Bellomo, L. Gibelli, A. Reali, On a unified multiscale vision of behavioral crowds. Math. Models Methods Appl. Sci. **30**(1), 1–22 (2020)
8. N. Bellomo, R. Bingham, M. Chaplain, G. Dosi, G. Forni, D. Knopoff, J. Lowengrub, R. Twarock, M.E. Virgillito, A multi-scale model of virus pandemic: heterogeneous interactive entities in a globally connected world. Math. Models Methods Appl. Sci. **30**, 1591–1651 (2020). Open Access
9. N. Bellomo, F. Brezzi, M. Chaplain, Modeling virus pandemics in a globally connected world a challenge towards a mathematics for living systems. Math. Models Methods Appl. Sci. **31** 2391–2397 (2021)
10. N. Bellomo, D. Burini, G. Dosi, L. Gibelli, D. A. Knopoff, N. Outada, P. Terna, M.E. Virgillito, What is life? A perspective of the mathematical kinetic theory of active particles. Math. Models Methods Appl. Sci. **31**, 1821–1866 (2021). Open Access
11. N. Bellomo, D. Burini, N. Outada, Pandemics of mutating virus and society: a multi-scale active particles approach. Philos. Trans. A R. Soc. **380**, 20210161 (2022)
12. A. Bellouquid, M. Delitala, *Modelling Complex Biological Systems - A Kinetic Theory Approach*. Series: Modeling and Simulation in Science, Engineering and Technology (Birkhäuser, Boston, 2006)
13. E. Callaway, Coronavirus vaccines: five key questions as trials begin. Nature **579**(7800), 481–481 (2020)
14. M. Cecconi, G. Forni, A. Mantovani, Ten things we learned about COVID-19. Intensive Care Med. **46**(8), 1590–1593 (2020)
15. E.L. Cooper, Evolution of immune system from self/not self to danger to artificial immune system. Phys. Life Rev. **7**, 55–78 (2010)
16. D. Cyranoski, Profile of a killer: the complex biology powering the coronavirus pandemic. Nature **581**(7806), 22–26 (2020)
17. G. Dosi, M. Piva, M.E. Virgillito, M. Vivarelli, Embodied and disembodied technological change: the sectoral patterns of job-creation and job-destruction. Res. Policy **50**(4), 104199 (2021)
18. R. Eftimie, L. Gibelli, A kinetic theory approach for modelling tumour and macrophages heterogeneity and plasticity during cancer progression. Math. Models Methods Appl. Sci. **30**(4), 659–683 (2020)
19. F. Flandoli, E. La Fauci, M. Riva, Multiscale aspects of virus dynamics in *Predicting Pandemics in a Globally Connected World, Vol. 1*, *Modeling and Simulation in Science, Engineering, and Technology*, Chap. X (Birkhäuser, New York, 2022)
20. J.F. Fontanari, Productivity in times of Covid-19: an agent-based model approach, in *Predicting Pandemics in a Globally Connected World, Vol. 1*, *Modeling and Simulation in Science, Engineering, and Technology*, Chap. X (Birkhäuser, New York, 2022)
21. Q. Gao, J. Zhuang, T. Wu, H. Shen, Transmission dynamics and quarantine control of COVID-19 in cluster community, in *Predicting Pandemics in a Globally Connected World, Vol. 1*, *Modeling and Simulation in Science, Engineering, and Technology*, Chap. X (Birkhäuser, New York, 2022)

22. D. Kim, A. Quaini, A kinetic theory approach to model pedestrian dynamics in bounded domains with obstacles. Kinet. Relat. Model. **12**, 1273–1296 (2019)
23. D. Kim, A. Quaini, Coupling kinetic theory approaches for pedestrian dynamics and disease contagion in a confined environment. Math. Models Methods Appl. Sci. **30**, 1893–1915 (2020)
24. D. Kim, A. Quaini, A 2D kinetic model for crowd dynamics with disease contagion, in *Predicting Pandemics in a Globally Connected World, Vol. 1*, *Modeling and Simulation in Science, Engineering, and Technology*, Chap. X (Birkhäuser, New York, 2022)
25. S.M. Kissler, C. Tedijanto, E. Goldstein, Y.H. Grad, M. Lipsitch, Projecting the transmission dynamics of SARS-CoV-2 through the post pandemic period. Science **368**(6493), 860–868 (2020)
26. N. Sfakianakis, A. Madzvamuse, M.A.J. Chaplain, A hybrid multiscale model for cancer invasion of the extracellular matrix. Multiscale Model. Simul. **18**(2), 824–850 (2020)
27. N. Vabret, Current state of the science. Immunity (2020). https://doi.org/10.1016/j.immuni.2020.05.002. Online ahead of print
28. M. Zagour, Multiscale derivation of a time-dependent SEIRD reaction-diffusion system for COVID-19, in *Predicting Pandemics in a Globally Connected World, Vol. 1*, *Modeling and Simulation in Science, Engineering, and Technology*, Chap. X (Birkhäuser, New York, 2022)

Understanding COVID-19 Epidemics: A Multi-Scale Modeling Approach

Maíra Aguiar, Vizda Anam, Nicole Cusimano, Damián Knopoff, and Nico Stollenwerk

1 Introduction

More than eighteen months have passed since a severe respiratory syndrome (COVID-19) caused by a new coronavirus (SARS-CoV-2) was identified in China [59] and spread rapidly around the globe. COVID-19 was declared a pandemic by the World Health Organization (WHO) in March, 2020 [24, 60]. As of the beginning of August, 2021, around 200 million cases were confirmed with more than 4 million deaths, a global case fatality ratio of approximately 2% [57, 61].

COVID-19 symptoms can range from asymptomatic/mild to severe illness, and disease severity and death occurring according to a hierarchy of risks [3], with age and pre-existing health conditions enhancing disease severity. With an unprecedented global health burden arising, the collective behavior of societies has been significantly affected by the extreme measures implemented to control disease transmission. Leading to serious socioeconomic problems, the COVID-19 pandemic is considered by the World Bank Global Economic Prospects to have caused the deepest global recession since the Second World War (WW II) [58].

Vaccines against COVID-19 have been developed in record time and are now globally distributed. With different efficacies, COVID-19 vaccines are remarkably effective against severe disease. However, the so-called sterilizing immunity, occurring when vaccinated individuals cannot transmit the virus, is still being evaluated. Four vaccines are now licensed for emergency use in Europe: two mRNA-type vaccines, Pfizer-BioNTech and Moderna, with about 95% vaccine efficacy after second dose, and two viral vector vaccines, the ones by Oxford-

M. Aguiar · V. Anam · N. Cusimano · D. Knopoff (✉) · N. Stollenwerk
Basque Center for Applied Mathematics, Bilbao, Spain
e-mail: maguiar@bcamath.org; vanam@bcamath.org; ncusimano@bcamath.org; dknopoff@bcamath.org; nstollenwerk@bcamath.org

N. Bellomo, M. A. J. Chaplain (eds.), *Predicting Pandemics in a Globally Connected World, Volume 1*, Modeling and Simulation in Science, Engineering and Technology, https://doi.org/10.1007/978-3-030-96562-4_2

AstraZeneca and Johnson & Johnson's Janssen, with about 70% vaccine efficacy upon full immunization, i.e., with two doses in the first case and one dose in the latter [14, 25, 28, 41, 45, 55]. Note that the above-mentioned vaccine efficacies are under constant evaluation, especially now that new SARS-CoV-2 variants have been identified.

In Spain, the new coronavirus infection was first notified on January 1st, 2020, and by March 13, cases had been confirmed in all 50 provinces of the country. A nationwide State of Alarm was declared on March 15, 2020, with a national lockdown becoming effective on March 16, 2020. All residents were mandated to remain in their normal residences except to purchase food and medicines, work, or attend emergencies. Lockdown restrictions also mandated the temporary closure of non-essential shops and businesses, including bars, restaurants, cafes, cinemas, and commercial and retail businesses. In the Basque Country, an autonomous community in northern Spain with 2.2 million inhabitants, the first cases of COVID-19 were notified on March 4, 2020. Ruled by the same Spanish decrees, lockdown measures were implemented accordingly and in time.

Research on mathematical modeling became imperative and very influential to understand the epidemiological dynamics of COVID-19 spreading and control over the course of the pandemic under different scenarios. An enormous quantity of epidemiological modeling peer-reviewed articles and preprints studying COVID-19 dynamics have appeared since the pandemic started. Aiming to predict the spread of the disease in a population, modeling task forces were created around the globe. Using, most of the time, simple models such as the SIR (Susceptible-Infected-Recovered) or SEIR (Susceptible-Exposed-Infected-Recovered) in mechanistic or probabilistic frameworks, researchers were requested to provide projections about specific disease-related variables such as hospitalizations, intensive care units admissions (ICUs), and deaths.

Already in March 2020, a multidisciplinary task force (the Basque Modeling Task Force [BMTF]) was created to assist the Basque health managers and Government during the COVID-19 responses. Within the BMTF, a stochastic SHARUCD modeling framework was developed [4, 9–11, 51]. As an extension of the simple SIR model, this flexible framework considers populations of susceptible individuals (S), severe cases prone to hospitalization (H), mild, sub-clinical, or asymptomatic (A), recovered (R), and patients admitted to intensive care units (U). The recorded cumulative positive cases, which includes all new positive cases for each class of H, A, U, R, are counted within the C classes, including the deceased (D) cases.

Able to describe the COVID-19 epidemic in terms of disease spreading, the SHARUCD model gives accurate projections (see Fig. 1) on hospitalizations, ICU admissions, and deceased cases from March 2020 to December 2020, when vaccination roll-out started. The modeling framework was used to monitor the COVID-19 epidemiological dynamics in the Basque Country while the lockdown measures were relaxed and tightened over time, evaluating also the impact of non-pharmaceutical interventions and social distancing.

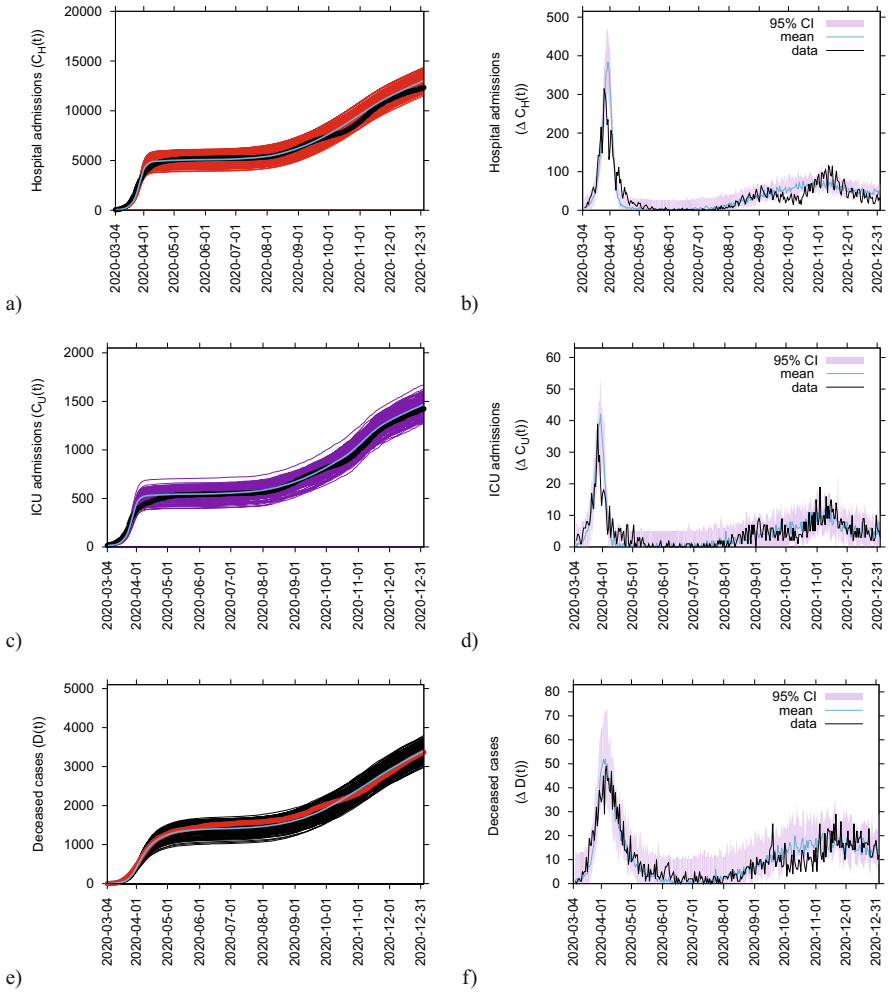


Fig. 1 From March 4 to December 31, 2020, on the left hand side we plot the ensemble of stochastic realizations of the SHARUCD model for cumulative cases. In (**a**) cumulative hospitalized cases $C_H(t)$, in (**c**) cumulative ICU admissions $C_U(t)$, and in (**e**) cumulative deceases cases $D(t)$. The mean of the stochastic realizations is plotted in light blue. Empirical data is plotted as black dots for hospitalizations and ICU admissions, and red dots for deceased cases. On the right hand side, we plot the model results for the daily incidences. In (**b**) daily hospitalized cases, in (**d**) daily ICU admissions, and in (**f**) daily deceased cases. Empirical data are plotted as a black line for all three cases, while the mean of 200 stochastic realizations is plotted in light blue. The 95% confidence intervals are obtained empirically from the stochastic realizations and are plotted as light purple shadow

It is worth stressing that in order to build useful models a close collaboration with field epidemiologists, laboratory researchers in virology, immunology, and biology,

as well as with public health stakeholders is needed [2, 5, 7, 8]. Moreover, a constant good data input is essential for model parameterization and validation [1, 6, 34, 43].

Modeling refinements were validated by epidemiological data continuously collected and provided by the Basque Health Department and the Basque Health Service (Osakidetza). Results on the evolution of the epidemic in the Basque Country are regularly updated and publicly available on the "SHARUCD Dashboard" [53].

The SHARUCD modeling framework and its refinements will be presented in the first part of this chapter. Basic concepts on vaccination toward herd immunity and the impact of vaccination roll-out, considering heterogeneity on vaccine efficacy for hospitalization and overall infection, will also be discussed.

Population heterogeneity is also an important feature to be considered when evaluating the effects of control measures applied to different segments of society. For example, pre-existing comorbidities and age are considered important factors affecting disease severity during the COVID-19 pandemic [3, 32] and, for that, the use of an age-structured population modeling approach would be appropriate to quantify the role of different age groups on disease spreading and to evaluate the impact of intervention measures for each population stratum (e.g., [12, 26, 27]).

Some useful contributions in this direction are provided by McBryde et al. [37] and Rozhnova et al. [44], for example, where age-structured models for SARS-CoV-2 fitted to hospital admission and seroprevalence data were used to estimate the impact of school contacts on transmission of the disease and to assess the effects of school-based measures, modeling interactions using mixing matrices [42].

There are, however, many different ways to include heterogeneity in dynamical models. In this chapter, we will also present an exploratory exercise using a modeling approach based on the kinetic theory of active particles (KTAP) published in [17]. The KTAP approach is very versatile since it allows the population to be subdivided into functional subsystems (FS) according to the problem under study. For instance, age structure or population with comorbidities can be easily introduced to stratify the population with a social network structure [12].

The KTAP approach also deals with the multiscale nature of the system under study. Besides discussion on contagion at the level of a population, where the virus is transmitted from infected to healthy individuals by short-range interactions, the complexity of the system increases with the inclusion of multiscale information, such as within-host features [23]. In this case, the micro-scale corresponds to virus particles and immune cells, which induce the dynamics at the higher scale of individuals who carry an infection, at meso-scale level, also affecting the collective behavior of individuals, which are then analyzed at the macro-scale level [17]; see illustration in Fig. 2.

This chapter is structured as follows. Section 2 presents the SHARUCD model and its refinements, both in the stochastic and deterministic versions, which are being successfully used to assist public health managers and policy makers in the Basque Country, Spain. Section 3 introduces the KTAP model for the spread of a disease. With an exploratory analysis, a possible way to deal with heterogeneity and multiscale features is presented. Section 4 concludes this work, with a discussion

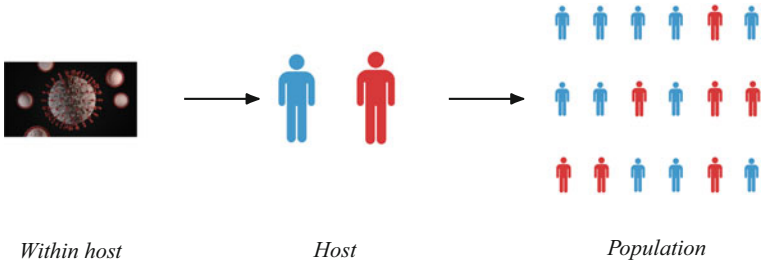


Fig. 2 Schematic representation of different scale dynamics. COVID-19 virus image on the left credit: Photo by Viktor Forgacs on Unsplash

on both models, presenting the connection of this research with the current epidemiological situation of vaccine impact and further research perspectives.

2 Mathematical Modeling Applied to Infectious Diseases: COVID-19 as a Case Study

Epidemiological models are formal frameworks to convey ideas about the components of a host–pathogen interaction and can be used as a tool to understand and predict the spread of infectious diseases as well as to evaluate the impact of control measures in different epidemiological scenarios.

Mathematical models were introduced into infectious disease epidemiology in the early twentieth century, and a series of deterministic compartmental models such as SIS (susceptible-infected-susceptible) and SIR (susceptible-infected-recovered) have been proposed based on the flow patterns between compartments of hosts in a population, which is divided into different subgroups for the considered disease-related stages.

One way to visualize these models is by using state-flow diagrams, see Fig. 3, where circles represent the compartments for each disease-related stage and arrows indicate the transitions for disease progression.

Disease propagation is an inherently stochastic phenomenon and there are a number of reasons why one should use stochastic models to capture the disease transmission process. However, the mean-field approximation, where the dynamics of the mean quantities are approximated by neglecting correlations, is often used as a good approximation to get a first understanding of the behavior of stochastic systems in certain parameter regions [40, 48, 49].

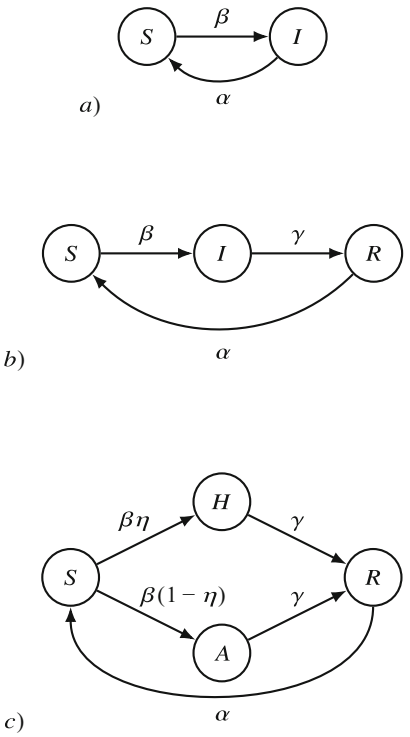


Fig. 3 State-flow diagram of simple epidemiological models. Disease-related stages are susceptibles S, infected I, and eventually stratified as asymptomatic A or hospitalized H, and recovered R. In (**a**) the SIS type model, in (**b**) the SIR type model, and in (**c**) the SHAR type model. For a host population of N individuals, with infection rate β, recovery rate γ, and waning immunity rate α. In the SHAR type model, η is the proportion of infected individuals prone to hospitalization, while $1-\eta$ will develop mild/asymptomatic infection

2.1 The SIR and SHAR Models

The SIR (susceptible-infected-recovered) model, see Fig. 3b, is one of the simplest compartmental models, dividing the observed population into three groups: the class S of susceptible individuals to the considered disease, the class of infected individuals I, and the class R of individuals who have recovered from the infection. In analogy with chemical reactions, the dynamics within the typical SIR framework with infection rate β, recovery rate γ, and waning immunity rate α can be illustrated by the scheme

$$\begin{array}{ccc} S+I & \xrightarrow{\beta} & I+I \\ I & \xrightarrow{\gamma} & R \\ R & \xrightarrow{\alpha} & S \end{array} \tag{1}$$

which translates into the following ODE system describing the temporal evolution of the number of individuals in each of the three model compartments:

$$\begin{aligned}\frac{dS}{dt} &= \alpha R - \beta \frac{S}{N} I,\\ \frac{dI}{dt} &= \beta \frac{S}{N} I - \gamma I,\\ \frac{dR}{dt} &= \gamma I - \alpha R,\end{aligned} \tag{2}$$

with N denoting the population size, i.e., $N = S + I + R$.

The stochastic SIR epidemic is modeled as a time-continuous Markov process to capture population noise. The temporal dynamics for the probability $p(S, I, t)$ of having an integer number S of susceptible and I of infected individuals at time t can be given as the following master equation [54]:

$$\begin{aligned}\frac{d}{dt} p(S, I, t) = &\frac{\beta}{N}(S+1)(I-1)\ p(S+1, I-1, t)\\ &+\gamma(I+1)p(S, I+1, t)\\ &+\alpha(N-(S-1)-I)\ p(S-1, I, t)\\ &-\left(\frac{\beta}{N} SI + \gamma I + \alpha(N-S-I)\right)\ p(S, I, t),\end{aligned} \tag{3}$$

while the number of recovered individuals follows from the constant population size assumption $R = N - (S + I)$. Letting $\mathbf{x}$ denote the state vector of densities, the master equation for the probabilities $p(\mathbf{x}, t)$ can be expressed in terms of n transitions $w_j(\mathbf{x})$ and small deviations from state $\mathbf{x}$ given by $\Delta\mathbf{x}_j$ as

$$\frac{d}{dt} p(\mathbf{x}, t) = \sum_{j=1}^{n} \left(N w_j(\mathbf{x} + \Delta\mathbf{x}) p(\mathbf{x} + \Delta\mathbf{x}_j, t) - N w_j(\mathbf{x}) p(\mathbf{x}, t) \right), \tag{4}$$

where $\Delta\mathbf{x}_j := \frac{1}{N}\mathbf{r}_j$ for suitable shifting vectors $\mathbf{r}_j$. Specifically, for the SIR model described above, we have $\mathbf{x} := (x_1, x_2)^T$ with $x_1 := S/N$ and $x_2 := I/N$, $n = 3$, and the following transitions $w_j(\mathbf{x})$ with corresponding shifting vectors $\mathbf{r}_j$:

$$\begin{aligned} w_1(\mathbf{x}) &= \beta x_1 x_2 &&, \mathbf{r}_1 = (1, -1)^T\\ w_2(\mathbf{x}) &= \gamma x_2 &&, \mathbf{r}_2 = (0, 1)^T\\ w_3(\mathbf{x}) &= \alpha(1 - x_1 - x_2) &&, \mathbf{r}_3 = (-1, 0)^T \quad . \end{aligned} \tag{5}$$

This process can be simulated by the Gillespie algorithm giving stochastic realizations of infected and susceptible individuals in time [30, 31].

To distinguish between mild and severely infected cases, the SIR framework can be extended into the so-called SHAR model, see Fig. 3c, where H stands for individuals developing a severe form of the disease and likely being hospitalized,

while A refers to infected individuals who are asymptomatic or have a mild form of the disease. This system includes two additional epidemiological parameters: η and ϕ. While the severity ratio η gives the fraction of infected individuals who develop severe symptoms (and hence $1-\eta$ gives the asymptomatic fraction of infections), the parameter ϕ is a scaling factor used to differentiate the infectivity $\phi\beta$ of mild/asymptomatic infections with respect to the baseline infectivity β of severe/hospitalized cases. The value of ϕ can be tuned to reflect different situations: a value of $\phi < 1$ reflects the fact that severe cases have larger infectivity than mild cases (e.g., due to enhanced coughing and sneezing), while $\phi > 1$ indicates that asymptomatic individuals and mild cases contribute more to the spread of the infection (e.g., due to their higher mobility and possibility of interaction) than the severe cases that are more likely to be detected and isolated [4].

In the case of COVID-19, we assume $\phi > 1$, since severe cases are likely hospitalized and isolated while mild/asymptomatic cases are often undetected and hence able to transmit the disease, contributing significantly more to the force of infection than the severe cases. The model dynamics can be illustrated by the following reaction scheme:

$$\begin{array}{rcl} S+H & \xrightarrow{\eta\beta} & H+H \\ S+H & \xrightarrow{(1-\eta)\beta} & A+H \\ S+A & \xrightarrow{\phi\eta\beta} & H+A \\ S+A & \xrightarrow{\phi(1-\eta)\beta} & A+A \\ S & \xrightarrow{\varrho} & A,H \\ H & \xrightarrow{\gamma} & R \\ A & \xrightarrow{\gamma} & R \\ R & \xrightarrow{\alpha} & S \end{array} \tag{6}$$

and the corresponding SHAR ODE system

$$\begin{aligned} \frac{dS}{dt} &= \alpha R - \beta\frac{S}{N}(H+\phi A+\varrho N), \\ \frac{dH}{dt} &= \eta\beta\frac{S}{N}(H+\phi A+\varrho N) - \gamma H, \\ \frac{dA}{dt} &= (1-\eta)\beta\frac{S}{N}(H+\phi A+\varrho N) - \gamma A, \\ \frac{dR}{dt} &= \gamma(H+A) - \alpha R, \end{aligned} \tag{7}$$

in which we have introduced also the import factor ϱ, which refers to the possibility of susceptible individuals becoming infected by an undetected infection chain, which started outside the studied population [35, 36].

The stochastic version of the presented SHAR model is obtained analogously to the basic SIR model shown above [4], and the dynamics for the probabilities $p(\mathbf{x}, t)$ can once again be given as in Eq. (4).

In particular, for the basic SHAR model with import ϱ and eventual waning immunity α (which matters, e.g., when new variants affect natural immunity of the host), $\mathbf{x}$ is the state vector of densities $x_1 := S/N, x_2 := H/N, x_3 := A/N$, and $x_4 := R/N$, while the $n = 5$ transitions and corresponding shifting vectors are given by:

$$\begin{aligned}
w_1(\mathbf{x}) &= \eta\beta x_1(x_2 + \phi x_3 + \varrho) &&, \mathbf{r}_1 = (1, -1, 0, 0)^T \\
w_2(\mathbf{x}) &= (1-\eta)\beta x_1(x_2 + \phi x_3 + \varrho) &&, \mathbf{r}_2 = (1, 0, -1, 0)^T \\
w_3(\mathbf{x}) &= \alpha(1 - x_1 - x_2 - x_3) &&, \mathbf{r}_3 = (-1, 0, 0, 1)^T \\
w_4(\mathbf{x}) &= \gamma x_2 &&, \mathbf{r}_4 = (0, 1, 0, -1)^T \\
w_5(\mathbf{x}) &= \gamma x_3 &&, \mathbf{r}_5 = (0, 0, 1, -1)^T \quad ,
\end{aligned} \tag{8}$$

(see, e.g., [4, 20, 29, 50, 51] for further details).

Figure 4 shows four snapshots in time (for $t = 10, 50, 60, 400$) of a single realization of the stochastic spatial SHAR model with import. Individuals are either susceptible (green), hospitalized (red), mild/asymptomatic infections (yellow), or recovered (blue). Assuming that the entire population (here $N = 10{,}000$) is initially susceptible at $t = 0$, the import term introduces the infection, producing isolated outbreaks of different sizes. While some of these outbreaks involve only a few individuals and quickly die out, with all individuals recovering, others may involve many more active infections (severe and/or mild and asymptomatic), spreading widely and eventually collapsing with neighboring clusters, leading to much larger outbreaks. An exponential growth of cases is to be expected when the community transmission is super-critical. A detailed analysis of spatial dynamics using the SHAR framework is ongoing and preliminary results will be briefly discussed in the last section of this chapter. More details on the role of import in epidemic models can be found in [11, 51].

2.2 The SHARUCD Modeling Framework

To describe the COVID-19 dynamics in the Basque Country, the basic SHAR model was extended by introducing the classes of Intensive Care Unit (ICU) admissions U and of deceased individuals D. Further, for comparison with the available cumulative empirical data, also the cumulative classes for hospitalized C_H, mild/asymptomatic infected C_A, ICU admitted C_U, and recovered C_R were included, counting all incoming cases in the dynamical compartments and

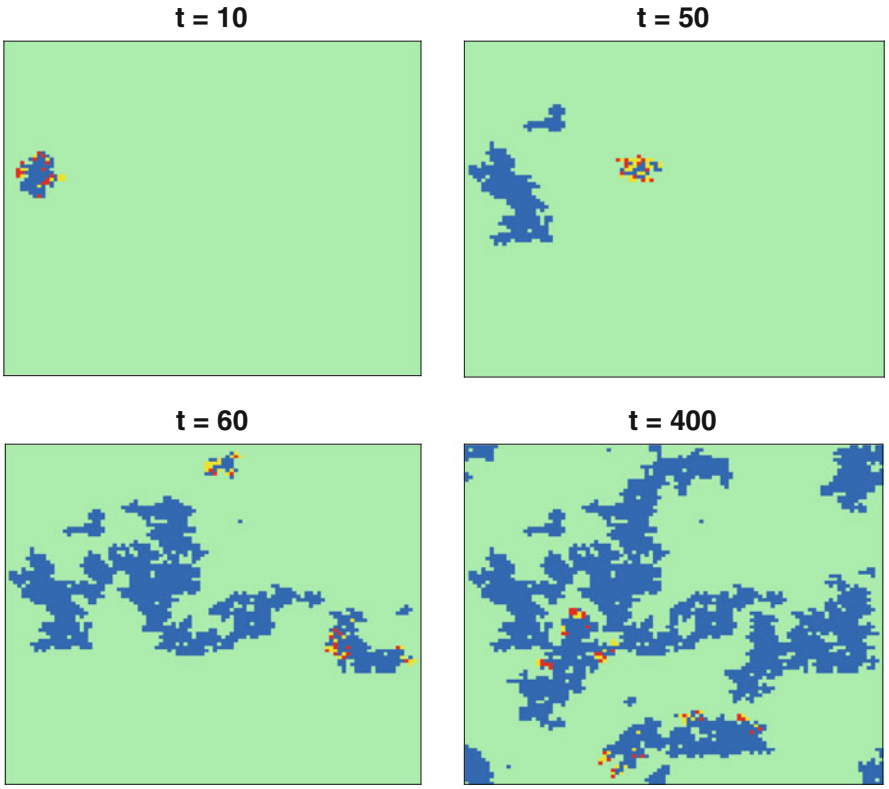


Fig. 4 Spatial configuration of a two-dimensional SHAR system with import at four different time points. The population of $N = 10{,}000$ is divided into susceptible (green), hospitalized (red), asymptomatic (yellow), and recovered (blue). Parameter values are: $\alpha = 0$, $\gamma = 1$, $\eta = 0.4$, $\beta = 0.85$, $\phi = 1.2$, and $\varrho = 10^{-9}$. Periodic boundary conditions are implemented

neglecting the outflows. A detection ratio ξ for mild/asymptomatic cases was also considered, since a proportion of mild/asymptomatic cases are detected by contact tracing/screening testing, and hence the number of positive tested infections is larger than the notified hospitalized cases.

In this model, disease severity is decided upon infection with a proportion η developing severe infection prone to hospitalization (and $1 - \eta$ developing mild/asymptomatic infection). Undetected asymptomatic cases are assumed to transmit the disease more efficiently ($\phi > 1$) than severe cases. Hospitalized individuals can recover, with a recovery rate γ, die, with a mortality rate μ, or go to an ICU facility, with an admission rate ν. Here, ICU admission is assumed to be a progression of disease severity after hospitalization.

The stochastic version of the basic SHARUCD model can be formulated through the master equation in the generic form of Eq. (4) with variables $x_1 := S/N$, $x_2 := H/N$, $x_3 := A/N$, $x_4 := R/N$, $x_5 := U/N$, $x_6 := C_H/N$, $x_7 := C_A/N$, $x_8 :=$

C_U/N, $x_9 := D/N$, and $x_{10} := C_R/N$. The state vector $\mathbf{x} := (x_1, ..., x_{10})^T$ gives the dynamics for the probabilities $p(\mathbf{x}, t)$, with $n = 10$ different transitions. The transitions $w_j(\mathbf{x})$ and the corresponding shifting vectors $\mathbf{r}_j$ are given by:

$$
\begin{aligned}
w_1(\mathbf{x}) &= \eta\beta x_1(x_2 + \phi x_3 + \varrho) && , \mathbf{r}_1 = (1, -1, 0, 0, 0, -1, 0, 0, 0, 0)^T \\
w_2(\mathbf{x}) &= \xi(1-\eta)\beta x_1(x_2 + \phi x_3 + \varrho) && , \mathbf{r}_2 = (1, 0, -1, 0, 0, 0, -1, 0, 0, 0)^T \\
w_3(\mathbf{x}) &= (1-\xi)(1-\eta)\beta x_1(x_2 + \phi x_3 + \varrho) && , \mathbf{r}_3 = (1, 0, -1, 0, 0, 0, 0, 0, 0, 0)^T \\
w_4(\mathbf{x}) &= \gamma x_2 && , \mathbf{r}_4 = (0, 1, 0, -1, 0, 0, 0, 0, 0, -1)^T \\
w_5(\mathbf{x}) &= (1-\xi)\gamma x_3 && , \mathbf{r}_5 = (0, 0, 1, -1, 0, 0, 0, 0, 0, 0)^T \\
w_6(\mathbf{x}) &= \gamma x_5 && , \mathbf{r}_6 = (0, 0, 0, -1, 1, 0, 0, 0, 0, -1)^T \\
w_7(\mathbf{x}) &= \nu x_2 && , \mathbf{r}_7 = (0, 1, 0, 0, -1, 0, 0, -1, 0, 0)^T \\
w_8(\mathbf{x}) &= \mu x_2 && , \mathbf{r}_8 = (0, 1, 0, 0, 0, 0, 0, 0, -1, 0)^T \\
w_9(\mathbf{x}) &= \mu x_5 && , \mathbf{r}_9 = (0, 0, 0, 0, 1, 0, 0, 0, -1, 0)^T \\
w_{10}(\mathbf{x}) &= \xi\gamma x_3 && , \mathbf{r}_{10} = (0, 0, 1, -1, 0, 0, 0, 0, 0, -1)^T \quad .
\end{aligned}
\tag{9}
$$

The mean-field ODE system corresponding to the above $w_j(\mathbf{x})$ and $\mathbf{r}_j$ is given in [10, 53] and was used to evaluate the model performance and its accuracy and to guide the modeling analysis.

The model was parameterized with empirical data provided by the Basque Health Department and the Basque Health Service (Osakidetza) for each disease-related variable. Parameters were estimated and fixed as the model was able to describe the disease incidence during the exponential phase of the outbreak; see Fig. 5.

The stochastic realizations of the model are calculated via the Gillespie algorithm [30, 31]. To investigate the parameter uncertainties, we calculate numerically the likelihood functions for each parameter conditioned on the others and the data, evaluating distances between simulations and data from all five variables, $D(t)$, $I_{cum}(t)$, $C_H(t)$, $C_U(t)$, and $C_R(t)$, for the exponential phase of the epidemic. The detailed analysis for the parameter estimation and uncertainties via likelihood functions can be found in [4, 10, 53].

Partial lockdown was implemented in the Basque Country on March 16, 2020, with effects observed on March 27 (see Fig. 5) well before the full lockdown of March 31, 2020 [10]. With the initial parameters estimated and fixed on the exponential phase of the epidemic, the next step was to model the effect of the disease control measures to describe the gradual slowing down of the epidemic.

2.3 Modeling the Implementation of Control Measures

The effect of the disease control measures was implemented by introducing for the infectivity parameter a smooth sigmoidal variation, which was able to describe well the gradual slowing down of the epidemic, reaching negative growth rate at the end of March, 2020. Specifically, the infection rate β became a time-dependent function $\beta(t)$ given by:

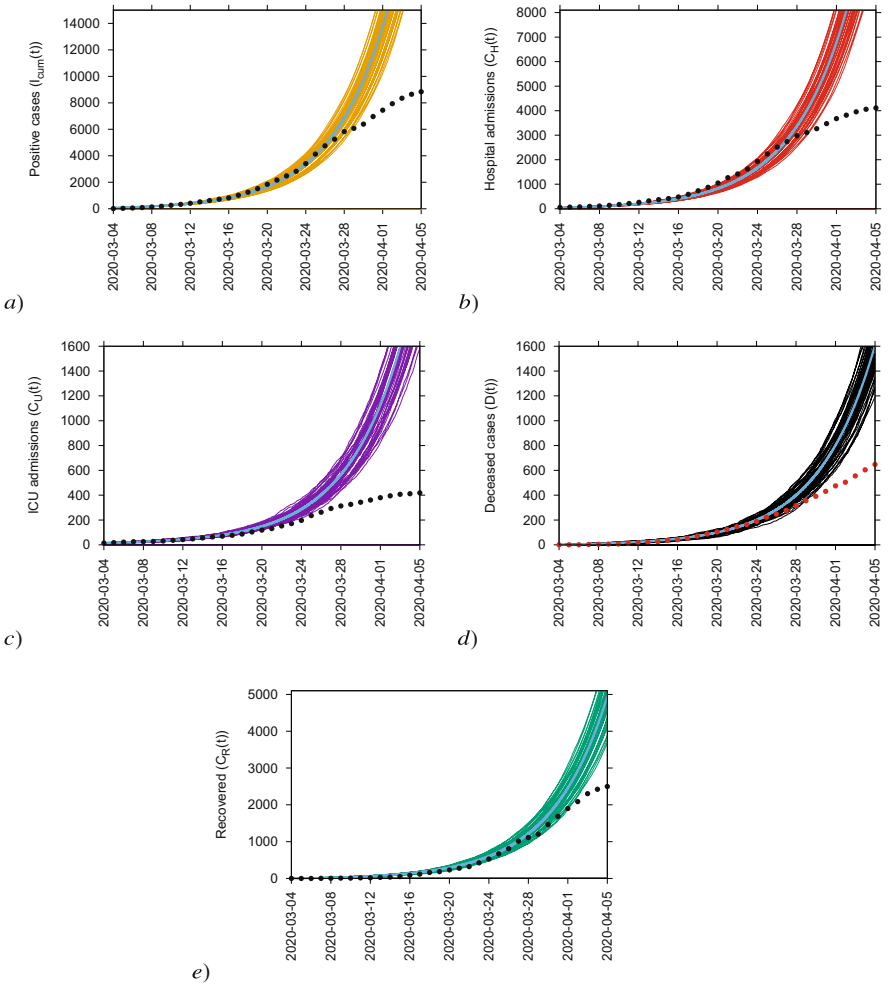


Fig. 5 From March 4 to April 4, 2020, ensemble of stochastic realizations of the baseline SHARUCD model. The mean-field solution is shown in light blue. In (**a**) cumulative detected positive cases $I_{cum}(t)$, in (**b**) cumulative hospitalized cases $C_H(t)$, in (**c**) cumulative ICU admissions $C_U(t)$, in d) cumulative deceased cases $D(t)$, and in e) cumulative recovered $C_R(t)$ (data on alive hospital discharges were used as a proxy for recovered individuals)

$$\beta(t) = \beta_0 \sigma_- \left(x(t)\right) + \beta_1 \sigma_+ \left(x(t)\right), \tag{10}$$

where $\sigma_-(x) = 1/(1+e^x)$ and $\sigma_+(x) = 1/(1+e^{-x})$ are downward and upward sigmoidal functions, respectively. The time-dependent function $x(t)$ is defined by $x(t) = a(t - t_c)$ and gives the timing and speed of the implementation of control

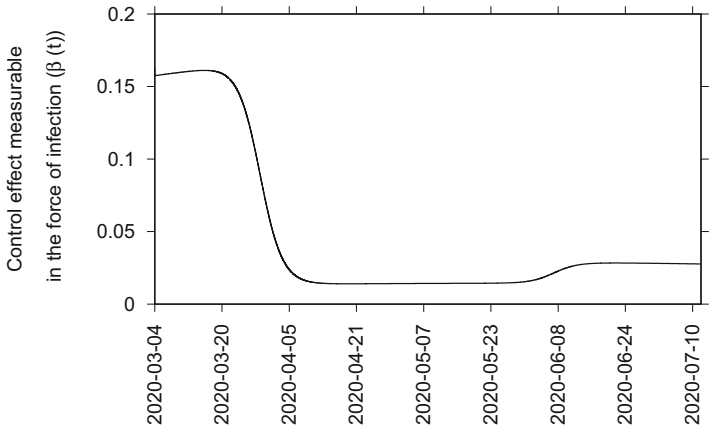


Fig. 6 Smooth variations in the infection rate $\beta = \beta(t)$ reproducing the effect of control measures, considering small seasonality. Parameter values are shown in Table 1. For the lockdown effect, $a = 0.38\ d^{-1}$ and $t_c = 25$ days after the initial time t_0 (corresponding here to March 4, 2020). For the lockdown lifting, $a_2 = 0.16\ d^{-1}$ and $t_{c_2} = 112$ days after t_0

measures. Further refinements with respect to Eq. (10), considering other smooth sigmoidal variations, have been implemented over the course of the pandemic, as control measures were introduced or relaxed (e.g., from March 4 to July 10, 2020, the profile of $\beta(t)$ is given in Fig. 6 with a smooth reduction of the transmission rate for the lockdown implementation and a later smooth increase corresponding to the gradual lockdown lifting). Low seasonality was also assumed to play a role, helping to keep transmission at low levels. For more details, please visit our SHARUCD Dashboard page [53].

Figure 7 shows an ensemble of stochastic realizations considering the effective control measures given by Eq. (10), up to mid-April, 2020. A good agreement was obtained for hospitalization, deceased cases, and recovered, with data lying in the median range of the 200 stochastic realizations. Notice that the deviation observed for the detected positive class I_{cum} is expected, due to the increased testing capacity started in March 22. The system was not adjusted to include the non-hospitalized cases. Moreover, the erratic behavior of the ICU admission cases remained to be investigated.

Using the data at hand, the momentary growth rates were calculated, see Fig. 8a, b, and compared with the momentary reproduction ratio, see Fig. 8c, confirming a negative growth of infection at the end of March, 2020 [8]. Two types of behavior were observed: positive cases, hospitalization, and, surprisingly, also the ICU admissions with the same sigmoidal decrease, shown in Fig. 8a, whereas the deceased and recovered were delayed for 8 and 10 days, respectively, shown in Fig. 8b.

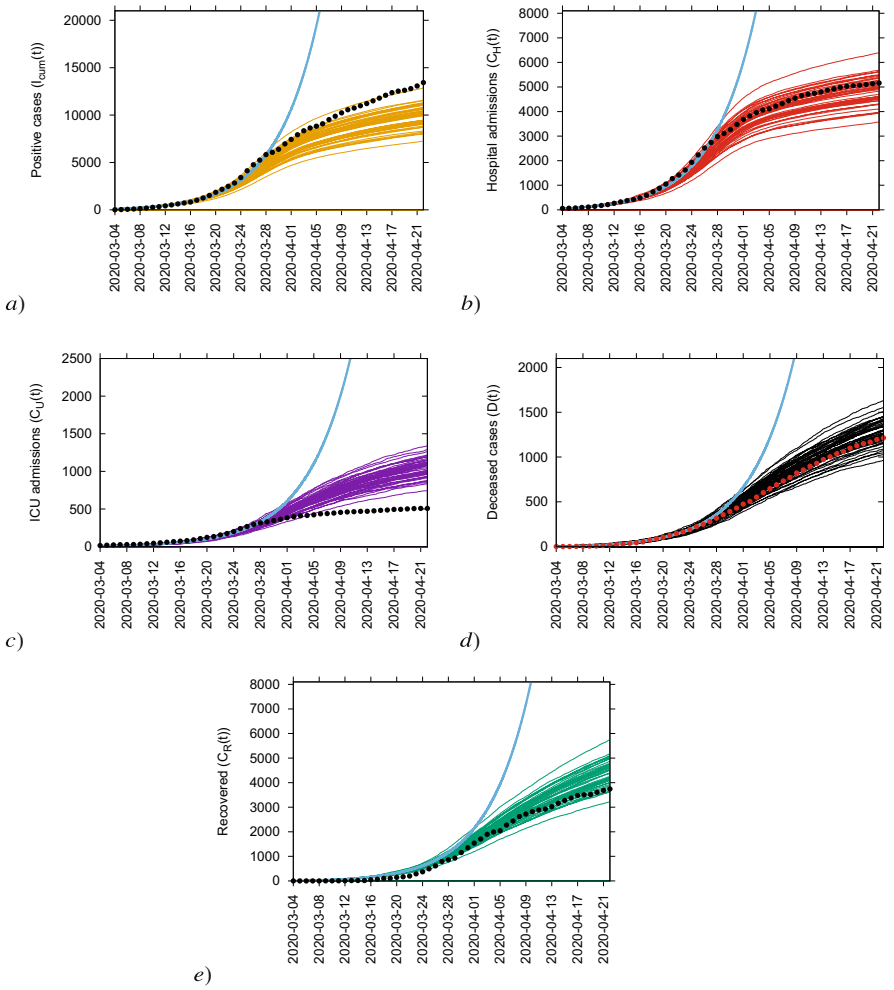


Fig. 7 From March 4 to April 21, 2020, ensemble of stochastic realizations of the baseline SHARUCD model. The mean-field solution without control is shown in light blue. In (**a**) cumulative detected positive cases $I_{cum}(t)$, in (**b**) cumulative hospitalized cases $C_H(t)$, in (**c**) cumulative ICU admissions $C_U(t)$, in (**d**) cumulative deceased cases $D(t)$, and in (**e**) cumulative recovered $C_R(t)$ (hospital discharges alive data were used as a proxy for recovered individuals)

This finding leads us to refine the baseline SHARUCD framework by changing the ICU admission rate ν into a ratio, with COVID-19 infection causing from asymptomatic up to very severe cases prone to hospitalization and to ICU admission.

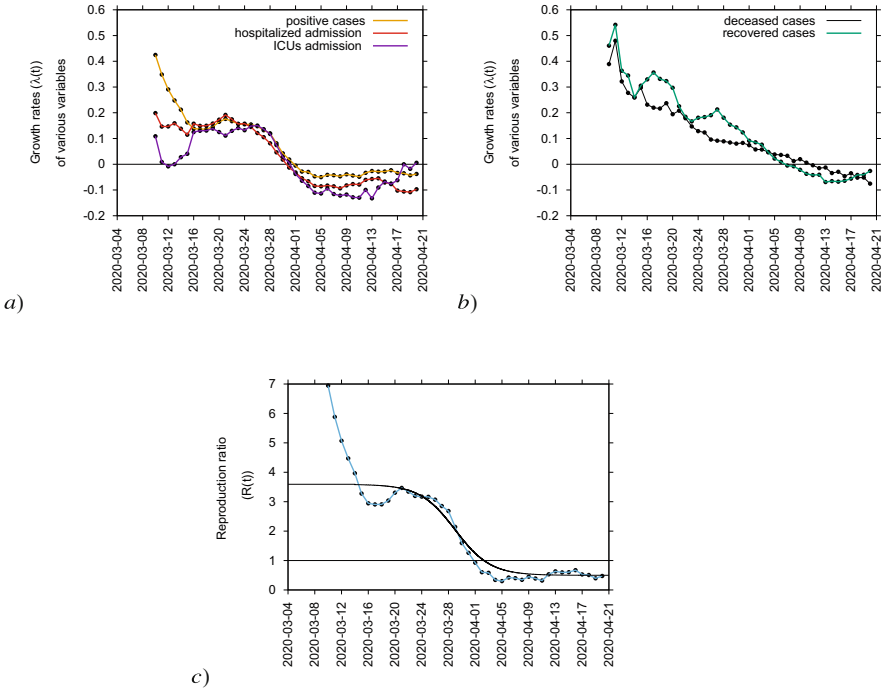


Fig. 8 From March 4 to April 21, 2020, the momentary growth rates for various variables show two types of behavior. In (**a**) detected positive cases in yellow, hospitalized cases in red, and ICU admissions in purple, cross the threshold on April 1, 2020. In (**b**) recovered cases in green and deceased cases in black, cross the threshold with a delay of 8 and 10 days, respectively. The momentary reproduction ratio is shown in (**c**)

2.4 The Refined SHARUCD Model

If previously it was assumed that hospitalized patients could either recover with recovery rate γ, be admitted to an ICU facility with rate ν, or die with disease-induced death rate μ, in light of the analysis of the momentary growth rates presented above, the baseline SHARUCD model was refined by considering ICU admissions to be a consequence of infection leading to severe disease, with a patient being immediately admitted to an ICU facility upon infection with ratio ν, analogously to the ratio η for hospitalizations. Note that although most of the hospitalizations were recovering or dying, some of those cases would eventually evolve into ICU cases and that was accounted for in the model as well.

The refined model is given by changes in the transition rates of the stochastic version, see [8], such that now the transition to ICU admission is synchronized with the transition to hospitalization and to the cumulative positive cases. Specifically, the transitions

$$\begin{aligned} w_1(\mathbf{x}) &= \eta\beta x_1(x_2 + \phi x_3 + \varrho) \ , \ \mathbf{r}_1 = (1, -1, 0, 0, 0, -1, 0, 0, 0, 0)^T \\ w_7(\mathbf{x}) &= \nu x_2 \qquad\qquad\qquad\quad \ , \ \mathbf{r}_7 = (0, 1, 0, 0, -1, 0, 0, -1, 0, 0)^T \end{aligned} \tag{11}$$

are changed into

$$\begin{aligned} w_1(\mathbf{x}) &= \eta(1-\nu)\beta x_1(x_2 + \phi x_3 + \varrho) \ , \ \mathbf{r}_1 = (1, -1, 0, 0, 0, -1, 0, 0, 0, 0)^T \\ w_7(\mathbf{x}) &= \eta\nu\beta x_1(x_2 + \phi x_3 + \varrho) \qquad \ , \ \mathbf{r}_7 = (1, 0, 0, 0, -1, -1, 0, -1, 0, 0)^T, \end{aligned} \tag{12}$$

with the parameter ν being adjusted from the ICU admission rate in units of d^{-1} into an ICU admission ratio $\nu \in [0, 1]$. The deterministic version of the model is hence given by:

$$\begin{aligned} \frac{d}{dt}S &= -\beta\frac{S}{N}(H + \phi A + \varrho N) \\ \frac{d}{dt}H &= \eta(1-\nu)\beta\frac{S}{N}(H + \phi A + \varrho N) - (\gamma + \mu)H \\ \frac{d}{dt}A &= (1-\eta)\beta\frac{S}{N}(H + \phi A + \varrho N) - \gamma A \\ \frac{d}{dt}R &= \gamma(H + U + A) \\ \frac{d}{dt}U &= \nu\eta\beta\frac{S}{N}(H + \phi A + \varrho N) - (\gamma + \mu)U \\ \frac{d}{dt}C_H &= \eta\beta\frac{S}{N}(H + \phi A + \varrho N) \\ \frac{d}{dt}C_A &= \xi \cdot (1-\eta)\beta\frac{S}{N}(H + \phi A + \varrho N) \\ \frac{d}{dt}C_R &= \gamma(H + U + \xi A) \\ \frac{d}{dt}C_U &= \nu\eta\beta\frac{S}{N}(H + \phi A + \varrho N) \\ \frac{d}{dt}D &= \mu(H + U) \end{aligned} \tag{13}$$

which is now able to describe the dynamics of the ICU admissions as well as the ones for hospitalized and deceased cases [10]; see Fig. 9.

Notice that the mean-field approximation of the stochastic system is given by the deterministic version in Eq. (13) (see [48]). For a complete analysis of the model, the reader is referred to [4]. Simulations of the stochastic SHARUCD model can be performed via the Gillespie algorithm [30, 31]. Figure 9 shows an ensemble of stochastic realizations for the COVID-19 epidemic in the Basque Country, from March 4 to June 16, 2020. The basic parameter values used for simulation of the model are shown in Table 1.

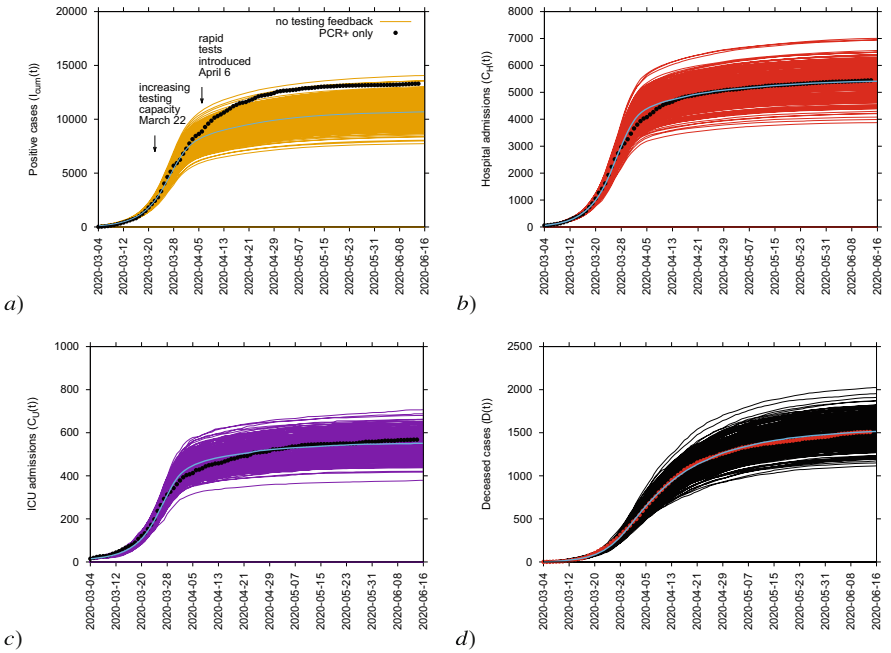


Fig. 9 From March 4 to June 16, 2020, ensemble of stochastic realizations of the refined SHARUCD model. The mean-field solution is shown in light blue. In (**a**) cumulative detected positive cases $I_{cum}(t)$,in (**b**) cumulative hospitalized cases $C_H(t)$, in (**c**) cumulative ICU admissions $C_U(t)$, and in (**d**) cumulative deceased cases $D(t)$

Table 1 SHARUCD model parameters

Model basic parameter	Description	Reference value
β	Infection rate	$3.25 \cdot \gamma$
γ	Recovery rate	$0.05\,\mathrm{d}^{-1}$
η	Proportion of hospitalization	0.075
ν	ICU admission ratio	0.09
ϕ	Ratio of mild/asymptomatic infections contributing to force of infection	1.65
ξ	Detection rate	[0.01–0.95]
μ	Disease-induced death rate	$0.02\,\mathrm{d}^{-1}$
ρ	Import	0.0006

2.4.1 Further Refinements: Detection Rate and Import

At the onset of the pandemic, data on COVID-19 infections were reflecting only severe cases requiring hospitalization, while the knowledge of asymptomatic and mild positive cases became available at a later stage, when testing capacity increased. This information heavily relies on local testing capacity and strategy

and is therefore highly variable in both time and space. To adjust to the observed deviation on the detected positive class I_{cum}, see Fig. 9a, due to the increased testing capacity over time, see Fig. 10a, the refined model considers a time-dependent detection ratio ξ assumed to be small in the introductory phase of the epidemic but later increased to a much higher level.

As testing capacity increased, the system was not able to describe quantitatively well the data for the recovered class R, since the hospital discharges data used as proxy for recovered individuals did not include the recovered individuals who were eventually tested positive but did not need hospitalization.

The model was also refined to describe isolated outbreaks after lockdown lifting, describing well the dynamics for positive detected cases, hospitalizations, ICU admission, and deceased cases from March 4 to December 31, 2020, shown in Fig. 1.

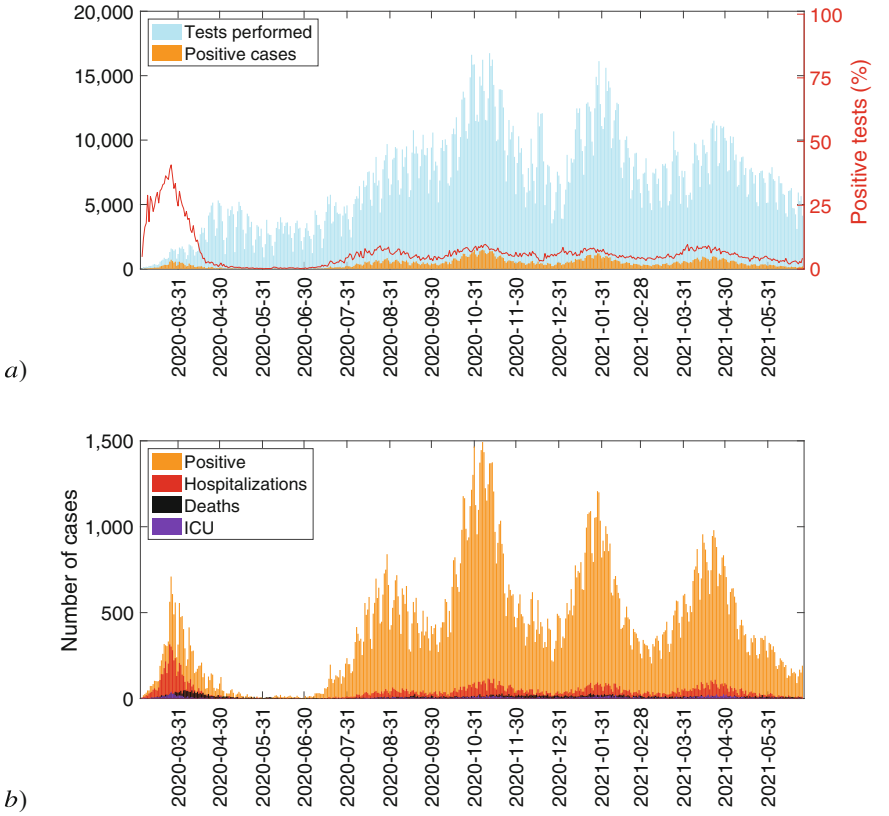


Fig. 10 From March 4 to June 24, 2021, in (**a**) COVID-19 tests performed in the Basque Country (light blue), positive PCR cases (yellow), and positivity rate (red line). In (**b**) COVID-19 detected cases in the Basque Country: Positive PCR cases are shown in yellow, hospitalizations in red, ICU admissions in purple, and deceased cases in black

We assume that an imported case is most likely a mobile asymptomatic infected individual, either a foreigner visiting the region or a local returning to the country without being detected by the current testing strategy, similarly to what one expects when country lockdowns are completely lifted and mobility is possible again. The so-called imported cases would be needed to describe the introductory phase of the epidemic, but we did not have any information on that at the start of exponential growth of the COVID-19 epidemic. The import factor becomes important again, after the full lockdown is lifted, allowing human mobility and asymptomatic disease transmission [11, 51]. For the present study, we assume ϱ to be much smaller than the other additive terms of the force of infection, given the strong observational insecurities on the data collected at the beginning of the outbreak.

Table 1 summarizes the starting parameter set used to describe the COVID-19 epidemic in the Basque Country from March 4, 2020. As it was previously mentioned in this chapter, model refinements were implemented during the course of the pandemic, with β, ξ, and ϱ becoming time-dependent parameters as non-pharmaceutical intervention actions were introduced and lockdown measures were implemented and relaxed over time. Therefore, it is important to mention that small adjustments on the presented parameter set were needed and with these small modifications we were able to match well data until December 31, 2020, as shown in Fig. 1. For further details, readers are referred to the SHARUCD Dashboard and references therein.

As a continuation of the BMTF efforts, the vaccination trial data for the vaccines that have been licensed for emergency use in Europe is evaluated. Results are implemented into the simple SHAR modeling framework and its extensions to get the qualitative overview of the impact of COVID-19 vaccination strategy in the Basque Country [52]. These results will be briefly described in the discussion section of this chapter and, for more details, readers are referred to [52].

3 KTAP Modeling Framework

The model presented in the previous section considers disease transmission at the population level. However, describing disease progression at the microscopic scale is also important and informative for the evaluation of pharmaceutical interventions for example, such as antiviral/antibiotics administration and vaccination. For instance, to investigate the infection process in a population of target cells, considering the immunological response of the host, the so-called within-host modeling framework is needed. Some recommended readings on this topic are [38, 39], and references therein, which introduce essential concepts on cell biology and immunology.

For respiratory diseases that cause damage to the lungs, like COVID-19, models should describe the dynamics of the viral load, which might lead to different asymptotic trends between full recovery and death by overload and even material corruption of the lung. A description of the dynamics of the lung in order to

detect those areas that are more susceptible to stretch overload in the pulmonary parenchyma is provided in [22]. Some useful recent contributions in this line are [33], which presents an interactive COVID-19 tissue simulator of viral dynamics of SARS-CoV-2 in a layer of epithelium, and [56] where a community-driven tissue simulator is developed.

Modeling ought to be developed within a multiscale approach. On the one hand, the dynamics of contagion must be treated at the macro-scale level of individuals and populations, while on the other the evaluation of the state of each individual (healthy, infectious, etc.) depending on the infection dynamics of the body cells as a result of the immunological response against the pathogen should be analyzed at the micro-scale level.

Both scales constantly interact and that is probably one of the most valuable virtues of the model presented in [17] and further refined in [12, 19], since the contagion at the macro-scale depends on the viral load of each individual, which in turn depends on the dynamics at the micro-scale. The model is based on the kinetic theory of active particles [18] and hereinafter will be referred to as the KTAP model.

The general framework supporting the KTAP modeling approach is defined by a selection of key features discussed in detail in the technical report [23]:

1. Individuals are viewed as *active particles* (a-particles), which are carriers of an internal state, called *activity*. In particular, the micro-state of every individual is described by a variable w corresponding to the level of activation of the immune defense. In addition, infected individuals are also characterized by a variable u representing the level of progression of the infection (ranging from mild to severe).
2. Contagion depends on the level of the infection as well as on an infection rate. The latter, as in the SHARUCD model, may be a time-dependent parameter that takes into account the implementation of specific health policies, like social distancing, lockdowns, etc.
3. The progression of the infection within each individual depends on the interaction between the pathogen replication and the immune response.

3.1 Modeling Contagion, Progression, and Recovery

Let us consider a population of N individuals homogeneously distributed in space. Each individual can belong, at time t, to one of the following compartments or functional subsystems (FS): susceptible (S-FS), infected (I-FS), recovered (R-FS), or deceased (D-FS). The S-FS is assumed to have only an outflow (into the I-FS), while R-FS and D-FS only have an inflow (from I-FS), i.e., we assume that recovered individuals get a long-lasting immunity and remain in that compartment.

The microscopic state of each individual entity is characterized by a variable $w \in [0, 1]$ describing the level of activation of the immune defense. It is worth stressing that in the original formulation of the model the value of w for a given individual

does not change over time. In order to simplify the identification of subgroups (or clusters of a-particles as in [21]) according to the individuals' immune response level (e.g., reflecting age or presence of comorbidities), the variable w is typically discretized and assumed to take values in the set

$$I_w = \{w_1 = 0, \ldots, w_k = \frac{k-1}{n-1}, \ldots, w_n = 1\}.$$

Within the I-FS class, individuals are also described by a variable $u \in [0, 1]$, which corresponds to the within-host progression of the pathogen invasion. Considering discrete values also for u in the set

$$I_u = \{u_1 = 0, \ldots, u_j = \frac{j-1}{m-1}, \ldots, u_m = 1\},$$

we have that $u_1 = 0$ describes the absence of the infection, while $u_j > 0$ indicates the presence of the disease with a variation of infection level. Increasing values of u toward 1 describe more aggressive states that may end up in the death of the host when reaching $u_m = 1$.

The number of immune levels n can be chosen by considering possible ways to disaggregate the population into age classes or morbidity groups, depending on the case under study (see, e.g., [3, 13]). On the other hand, m shall take into account the number of stages of a given disease after the initial entry of the pathogen into the host. In order to have all three compartments I, D, and R in the modeling framework, it is assumed that $m \geq 3$.

The dynamics is such that susceptible individuals, characterized by their microstate w_k, may become infected after an interaction with an infected a-particle. Within the I-FS, an a-particle is characterized by the pair (u_j, w_k), with $2 \leq j \leq m-1$. A competitive interaction between the pathogen and the immune system then starts and the transition into R-FS or D-FS depends on the result of this dynamics: namely, if the state u_1 (resp., u_m) is reached, the individual undergoes a transition into the recovered (resp., deceased) compartment. This is illustrated in Fig. 11.

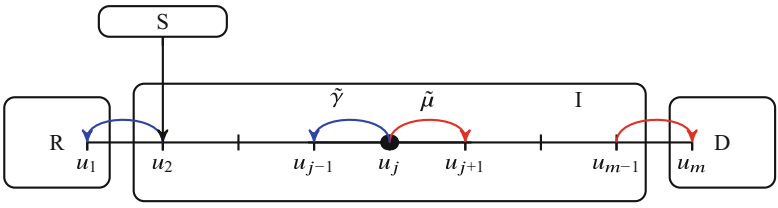


Fig. 11 Illustration of the transitions: susceptible individuals can get infected with an infection rate $\tilde{\beta}$, entering into the infected compartment (wide box in the middle) with state u_2. Then, competitive interactions between the pathogen that replicates with rate $\tilde{\mu}$ toward more aggressive infection states and the immune system, which acts with rate $\tilde{\gamma}$, with resulting transition into the R or D compartment

The representation of the system is given by the following distribution functions:

- $f_S^k(t)$ is the probability to find, at time t, a susceptible individual with micro-state w_k. The susceptible population at time t, $f_S(t)$, is simply computed as the sum $\sum_{k=1}^{n} f_S^k(t)$.
- $f_I^{j,k}(t)$ is the probability to find, at time t, an infected individual with micro-state (u_j, w_k). The prevalence of infection, at time t, is given by $f_I(t) = \sum_{k=1}^{n} \sum_{j=2}^{m-1} f_I^{j,k}$.
- $f_R^k(t)$ is the probability to find, at time t, a recovered individual with micro-state w_k. The cumulative recovered population $f_R(t)$ is simply computed as the sum $\sum_{k=1}^{n} f_R^k(t)$.
- $f_D^k(t)$ is the probability to find, at time t, a deceased individual with micro-state w_k. As for the recovered, the cumulative deceased population $f_D(t)$ is given by $\sum_{k=1}^{n} f_D^k(t)$.

Notice that $f_S(t) + f_I(t) + f_R(t) + f_D(t) = 1$ for all t and that the actual number of individuals belonging to each FS can be computed multiplying the distribution function by the total population N.

The dynamics of the system is described by using tools of the mathematical structures of the kinetic theory of active particles, see, e.g., Chapter 5 in [16]. The system of equations representing the evolution of the distribution functions, whose derivation can be followed in details in [17], is given by:

$$\begin{cases} \dfrac{d}{dt} f_S^k(t) = -\tilde{\beta} \displaystyle\sum_{s=1}^{n} \sum_{j=2}^{m-1} u_j \, f_S^k(t) \, f_I^{j,s}(t), \\ \dfrac{d}{dt} f_I^{j,k}(t) = \tilde{\beta} \, \delta_{2j} \left(\displaystyle\sum_{s=1}^{n} \sum_{p=2}^{m-1} u_p \, f_S^k(t) \, f_I^{p,s}(t) \right) + \tilde{\mu} \, u_{j-1} \, f_I^{j-1,k}(t) \\ \qquad + \tilde{\gamma} \, w_k \, f_I^{j+1,k}(t) - \tilde{\mu} \, u_j \, f_I^{j,k}(t) - \tilde{\gamma} \, w_k \, f_I^{j,k}(t), \\ \dfrac{d}{dt} f_R^k(t) = \tilde{\gamma} \displaystyle\sum_{k=1}^{n} w_k \, f_I^{2,k}(t), \\ \dfrac{d}{dt} f_D^k(t) = \tilde{\mu} \, u_{m-1} \displaystyle\sum_{k=1}^{n} f_I^{m-1,k}(t), \end{cases} \tag{14}$$

where for the first equation we have $k = 1, \ldots, n$, while for the second one $j = 2, \ldots, m-1$ and $k = 1, \ldots, n$. The first equation in (14) describes the infection of susceptible individuals due to interactions with infected ones. The second equation in (14) describes the dynamics within the infected population. The factor δ_{2j} denotes a Kronecker delta, meaning that the entry state upon infection is u_2. From that point, a competitive interaction between the pathogen and the immune

system starts. Finally, the third and fourth equations give the inflows into recovered and deceased classes, respectively, as a result of the aforementioned competitive interactions. Regarding model parameters, $\tilde{\beta}$ is the infection rate, $\tilde{\mu}$ is the pathogen progression rate, and $\tilde{\gamma}$ is the immune action rate toward recovery.

An extension of the system in Eq. (14) to networks can be found in [12].

3.2 Application of the KTAP Model to Selected Case Studies

The KTAP model introduced above has been applied to a variety of exploratory experiments aiming to understand the role of population heterogeneity in the propagation of the disease with an insight in non-pharmaceutical interventions (NPI) [17] and vaccination [19]. In this section, we present some selected case studies focusing on the intensity and timing of NPI and on the heterogeneity of the population, represented by the micro-states w_k.

In the following, let us consider a population of $N = 2.2$ million individuals (a-particles in this context). At time $t = 0$, almost all of them belong to the S-FS, with 100 a-particles in the I-FS, while R-FS and D-FS are empty. We consider $n = 5$ classes, in agreement with a possible choice among demographers and epidemiologists to disaggregate a population into age groups, but other possibilities can also be considered [13]. Although n and m need not be the same, here we also set $m = 5$.

3.2.1 Effect of Lockdown Measures and Restrictions Lifting

Let us first study the effect of lockdown measures, in order to understand how the amplitude and timing of the action influences the overall dynamics. The lockdown implementation is modeled by a reduction of the infection rate $\tilde{\beta}$ during a given time interval.

Figure 12 shows the dynamics of cumulative infected cases up to a given T_{max} for three different scenarios. In all three cases, the pathogen replication rate and the immune action rate are kept constant and equal to $\tilde{\mu} = 0.008$ and $\tilde{\gamma} = 0.03$, respectively.

In Fig. 12a the infection rate $\tilde{\beta}$ is initially equal to 0.6. At locking time $T_l = 25$, the effect of a lockdown is modeled by reducing the transmission rate $\tilde{\beta}$ to 0.06 until a reopening time $T_o = 50$ at which $\tilde{\beta}$ is increased to β_o. The curves corresponding to three different values of β_o are shown in the plot, up to $T_{max} = 100$.

Figure 12b shows a similar situation in which a lockdown is implemented, reducing the transmission rate $\tilde{\beta}$ from 0.6 to 0.06 at different locking times $T_l = 25, 30, 40$, giving an insight on the importance of intervention measures timing.

In Fig. 12c we consider three different opening times $T_o = 50, 70, 90$ at which the infection rate $\tilde{\beta}$ is increased back from 0.06 to 0.6. Notice that the three curves increase toward large values approaching N, showing that a late reopening

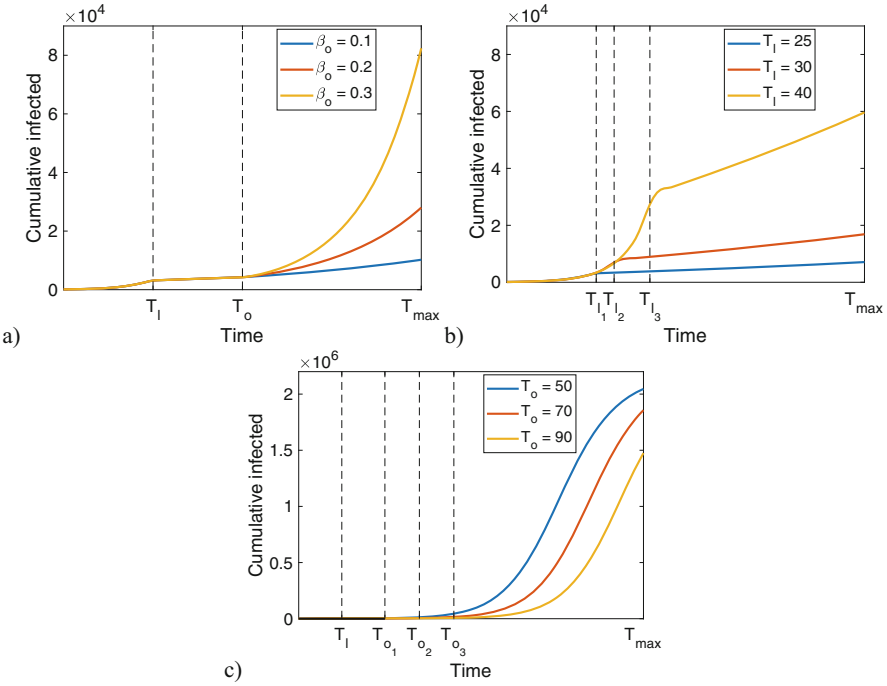


Fig. 12 Influence of lockdown measures and restrictions lifting. In (**a**) the transmission rate $\tilde{\beta}$ is reduced from 0.6 to 0.06 at locking time $T_l = 25$ and then increased again, at opening time $T_o = 50$, to different values β_o. In (**b**) the lockdown is implemented at three different locking times, with reduction of $\tilde{\beta}$ from 0.6 to 0.06. In (**c**) $\tilde{\beta}$ is increased back to 0.6 at three different opening times T_o. The other parameter values are $n = m = 5$, $\tilde{\gamma} = 0.03$, and $\tilde{\mu} = 0.008$

without control leads to a delayed explosion of infections. Here, $T_{max} = 200$ for visualization purposes.

3.2.2 Effect of Heterogeneity

Let us now study the effect of heterogeneity in the population by considering different distributions of immunity, which is a proxy for age and presence of other comorbidities.

The left panels of Fig. 13 show three possible immunity distributions: a centered symmetric distribution, describing a population with an average immunity level, a distribution that is skewed to the right, describing a population having a stronger immune system (e.g., a younger population), and one that is skewed to the left, describing a population having a weaker immune system (e.g., a vulnerable population composed by elderly individuals). In [12] these distributions are used to describe, respectively, populations in households, schools, and nursing homes.

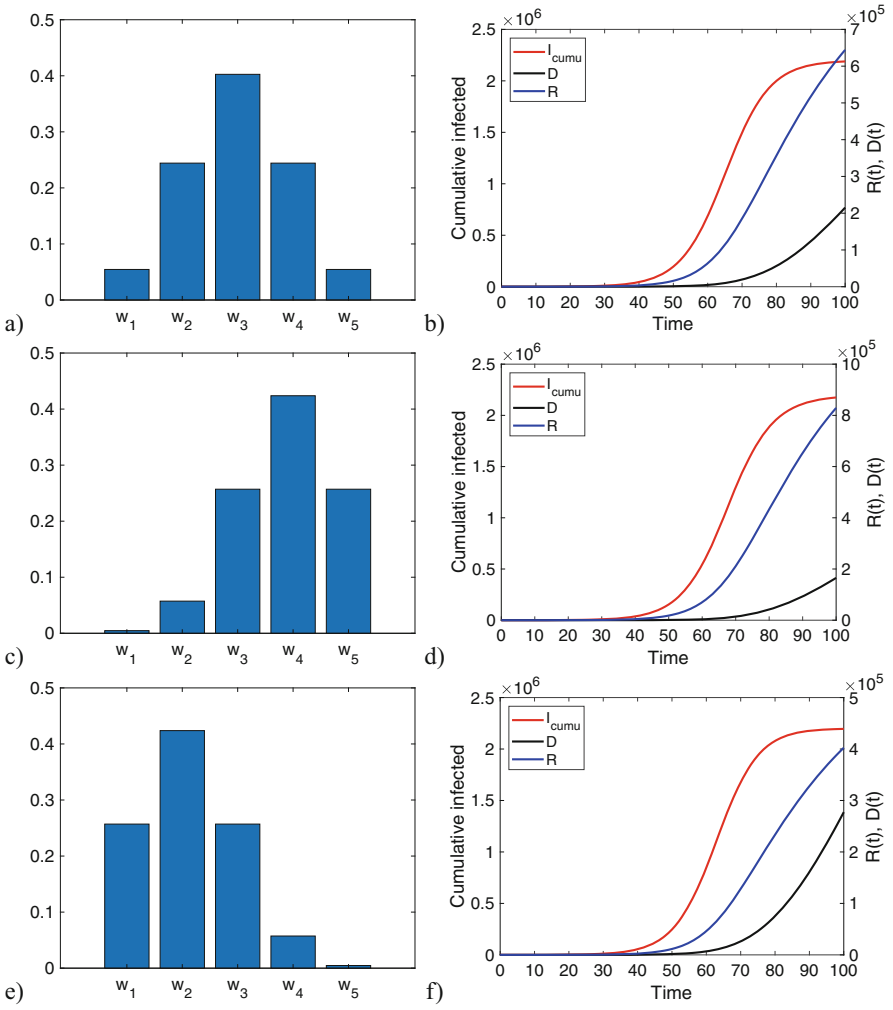


Fig. 13 The left panel shows the frequency distribution of immunity levels w_k in three cases: (**a**) centered, (**c**) skewed right, and (**e**) skewed left. On the right panel, we show the cumulative infected (left axis), recovered, and deceased cases (right axis) for the corresponding distributions on the left. No interventions are assumed in this case and the infection rate is fixed at $\tilde{\beta} = 0.6$

On the right panel of Fig. 13 we can see the dynamics for cumulative infected (left axis), recovered, and deceased cases (right axis) for the initial exponential phase for the three distributions of w_k shown on the left. The set of parameters is the same as above (with $\tilde{\beta} = 0.6$ unchanged during the considered time span), while the initial conditions reflect the distribution of w_k for each case. Notice that the immunity distribution actually influences the number of infected, recovered, and

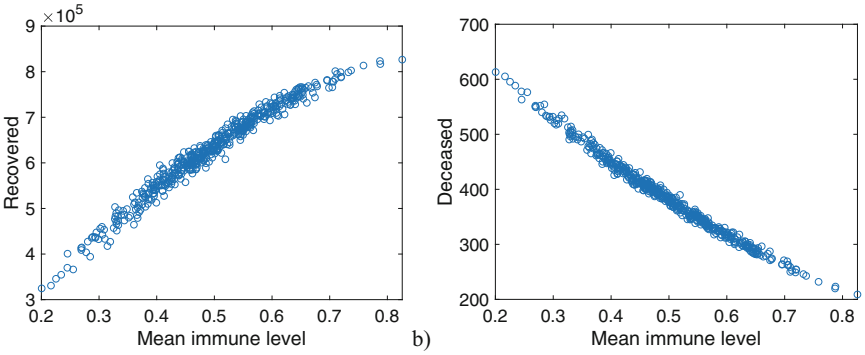


Fig. 14 Recovered and deceased cases at final time for different mean immune levels. The immunity distribution is chosen randomly for each of the 500 realizations of the experiment

deceased cases, since larger (resp., lower) values of w_k lead, eventually, to a faster recovery (resp., decease).

Finally, we perform an exercise to study the correlation between population heterogeneity and infection. Let us define the mean immune level by $\sum_{k=1}^{n} w_k f^k$, where f^k is the probability to find an a-particle with state w_k. In our case, $f^k = f_S^k + \sum_{j=1}^{m} f_I^{j,k} + f_R^k + f_D^k$ does not change with time due to the assumption that particles keep their micro-state w_k.

The scatterplot in Fig. 14 shows a strong positive (resp., negative) correlation between the mean immune level and the recovered (resp., deceased) for 500 realizations of the following experiment: the distribution of w_k is randomly chosen to define the initial conditions and the mean immune level is computed, then the KTAP model is run with the parameters $\tilde{\beta} = 0.6, \tilde{\gamma} = 0.03, \tilde{\mu} = 0.008$ up to $T_{max} = 100$. A marker represents the recovered and deceased cases at T_{max} for that particular immune level. This result emphasizes the role of heterogeneity and will be further discussed in the next section in the context of vaccination strategies.

4 Discussion

In the previous sections we have presented two different approaches to model the spread of an epidemic. The SHARUCD modeling framework introduced in Sect. 2 was developed within the Basque Modeling Task Force. As an extension of the basic SIR-type epidemiological models, the SHARUCD model stratifies the class of infected individuals into severe/hospitalized cases H and mild/asymptomatic A and includes further classes for intensive care unit admissions U and deceased D. The stochastic SHARUCD model can be regarded as a predictive model, which successfully describes the COVID-19 epidemic in the Basque Country in terms of disease spreading and control, as shown in Fig. 1. The model is currently used

to monitor COVID-19 transmission in the Basque Country (the complete model is described in detail in [10, 53]) and was able to provide accurate projections on the regional health system's necessities during the first wave of the pandemic and beyond. As explained in Sect. 2, the SHARUCD model was refined to analyze isolated outbreaks [11, 51], including now import to asymptomatic infection, after lifting of lockdowns, and increased detection of asymptomatic due to increasing testing capacity.

Investigation on critical fluctuations around the epidemiological threshold has shown that the lockdown measures implemented in the autumn 2020 were able to drive the growth of COVID-19 cases to the so-called sub-critical regime of community transmission, leading to a low notification of severe cases, up to the end of October 2020, and kept in stationarity even when the overall transmission rate increased by an enhanced mobility. The same behavior was also observed in different European regions [11, 51].

This study is a baseline for the understanding of the impact of the current vaccination programs around the globe giving insights on the role of community transmission and mobility regarding the reduction of severe disease. The system can be also evaluated under different vaccine efficacies and coverage, while population immunity is acquired by natural infection and vaccination, until finally reaching the so-called herd immunity threshold. The SHARUCD framework is under refinement to include the uneven vaccination roll-out strategy currently in place worldwide with preliminary results described in [52].

We use the epidemiological SHAR model framework to evaluate the effects of vaccination in different epidemiological scenarios of coverage and efficacy. Two vaccination models (one protecting against severe disease and the other protecting against infection as well as severe disease) are compared to evaluate the reduction of overall infections and hospitalizations. Differences in vaccine efficacy and coverage must be also considered for a more accurate evaluation of vaccine performance. Our results show that in an intermediate scenario of low to medium vaccination coverage and limited vaccine efficacy, vaccine performance will be influenced by the transmission level of mild and asymptomatic cases [52]. These results and concepts are of use to study also the epidemiological situation of vaccine waning immunity and immune escape by new variants.

On the other hand, the KTAP approach presented in Sect. 4 has so far been applied with exploratory purposes to model an infectious disease where a pathogenic agent spreads among a population of target cells, trying to evade the immune response. In this sense, the model is quite simple not even considering the nature of target cells (e.g., mainly gut and lung cells in the case of SARS-CoV-2 infection [15, 46]). However, this modeling approach can be adapted to explore different scenarios considering multi-scale features and heterogeneity of the population by stratifying it, e.g., according to age or comorbidities. For example, in [12] the population is distributed over a network where each node has a different social structure type: school, household, working place, and nursing home. Interactions can be modeled according to this stratification using, for example, mixing matrices [21, 37].

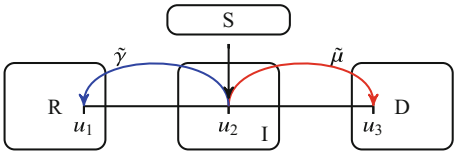


Fig. 15 Schematic representation of the simplified KTAP model with $n = 1$ and $m = 3$

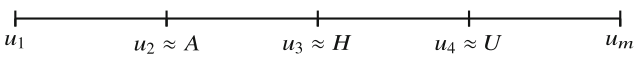


Fig. 16 Different severity levels of the disease

It is worth stressing that, if no heterogeneity is considered in the KTAP model (meaning $n = 1$) and only three levels of infection severity are considered (i.e., $m = 3$), we get, as a particular case, the SIR-type model configuration shown in Fig. 15. In this case, we can relate the dependent variables from both models as: $f_S(t) \approx x_1(t)$, $f_I(t) \approx x_2(t) + x_3(t) + x_5(t)$, $f_R(t) \approx x_4(t)$, and $f_D(t) \approx x_9(t)$ (see Eq. (9)).

A subsequent step is to relate the viral replication in the KTAP approach with the different compartments in the SHARUCD model, schematized in Fig. 16, and to properly model transitions not only between adjacent states but also to distant ones (for instance, a patient who has been immediately admitted to ICU facility upon infection). This perspective gives also an insight into a possible further stratification in the SHARUCD model, considering, e.g., age groups or accounting for subpopulations with comorbidities.

Finally, as vaccination against COVID-19 proceeds, both models introduced in this present chapter are under refinement to investigate the impact of population immunity by vaccination.

The KTAP approach has also been applied to conduct an exploratory analysis on the impact of vaccination targeting a given population. In [12] the effect of vaccination was modeled by "moving" some individuals with low w values to the highest level w_n. When the number of vaccinated particles increases, a strong reduction in terms of infected and deceased cases is observed, with a large negative correlation between the variables. This result confirms the relationship shown in Fig. 14 and the importance of immunization. In order to take into account the dynamics of the rollout of vaccines, ongoing research assumes the micro-states w_k to be time dependent.

Adding a spatial component to the models considered above is another way to introduce an additional scale into the modeling framework. Spatial dynamics is required in certain settings, e.g., when wanting to study differences in disease transmission dynamics depending on geographical location or when looking at the combined impact on transmission of different local/regional public health policies.

In Sect. 2 of this chapter, we have illustrated the effect of an import term in the stochastic and spatially dependent SHAR model. While in the non-spatial context determining the critical threshold of community transmission can be done analytically, in the spatially dependent case this is no longer possible and one has

to measure such threshold numerically instead. In general, a deviation from the corresponding mean-field approximation is to be expected since spatial correlations alter the results.

This type of analysis for the models presented in this chapter is ongoing and will be discussed in detail in our forthcoming publications.

Acknowledgments We thank the huge efforts of the whole COVID-19 Basque Modeling Task Force (BMTF), specially to Eduardo Millán, for collecting and preparing extended data sets on COVID-19 in the Basque Country, Javier Mar, Biodonostia Health Research Institute, and Joseba Bidaurrazaga Van-Dierdonck, Basque Health Department, for fruitful discussions on our modeling results. We thank Adolfo Morais Ezquerro, Vice Minister of Universities and Research of the Basque Government, for starting the BMTF initiative and for fruitful discussions. This research is supported by the Basque Government through the "Mathematical Modeling Applied to Health" Project, BERC 2018–2021 program, and by the Spanish Ministry of Sciences, Innovation and Universities: BCAM Severo Ochoa accreditation SEV-2017-0718. M. A. has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement no. 792494.

References

1. M. Aguiar, N. Stollenwerk, Dengvaxia efficacy dependency on serostatus: a closer look at more recent data. Clin. Infect. Dis. **66**(4), 641–642 (2018)
2. M. Aguiar, N. Stollenwerk, Dengvaxia: age as surrogate for serostatus. Lancet Infect. Dis. **18**(3), 245 (2018)
3. M. Aguiar, N. Stollenwerk, Condition-specific mortality risk can explain differences in COVID-19 case fatality ratios around the globe. Publ. Health **188**, 18–20 (2020)
4. M. Aguiar, N. Stollenwerk, SHAR and effective SIR models: from dengue fever toy models to a COVID-19 fully parametrized SHARUCD framework. Commun. Biomath. Sci. **3**(1), 60–89 (2020)
5. M. Aguiar, N. Stollenwerk, The impact of serotype cross-protection on vaccine trials: DENVax as a case study. Vaccines **8**, 674 (2020)
6. M. Aguiar, R. Paul, A. Sakuntabhai, N. Stollenwerk, Are we modeling the correct data set? Minimizing false predictions for dengue fever in Thailand. Epidemiol. Infect. **142**, 2447–59 (2014)
7. M. Aguiar, G.E. Coelho, L. Mateus, R. Rocha, J.E.M. Pessanha, L. Mateus, N. Stollenwerk, Dengue transmission during the 2014 FIFA World Cup in Brazil. Lancet Infect. Dis. **10**, 765–766 (2015)
8. M. Aguiar, N. Stollenwerk, S.B. Halstead, Modeling the impact of the newly licensed dengue vaccine in endemic countries. PLoS Neglect. Trop. D. **10**(12), e0005179 (2016)
9. M. Aguiar, J. Bidaurrazaga Van-Dierdonck, N. Stollenwerk, Reproduction ratio and growth rates: measures for an unfolding pandemic. PLoS ONE **15**, e0236620 (2020)
10. M. Aguiar, E. Millán Ortuondo, J. Bidaurrazaga Van-Dierdonck, J. Mar, N. Stollenwek, Modelling COVID 19 in the Basque Country from introduction to control measure response. Sci. Rep. **10**, 17306 (2020)
11. M. Aguiar, J. Bidaurrazaga Van-Dierdonck, J. Mar, N. Cusimano, D. Knopoff, V. Anam, N. Stollenwek, Critical fluctuations in epidemic models explain COVID-19 post-lockdown dynamics. Sci. Rep. **11**, 13839 (2021)

12. M. Aguiar, G. Dosi, D. Knopoff, M.E. Virgillito, A multiscale network-based model of contagion dynamics: heterogeneity, spatial distancing and vaccination. Math. Models Methods Appl. Sci. **31**, 2425–2454 (2021)
13. O.B. Ahmad, et al., Age standardization of rates: a new WHO standard. GPE Discussion Paper Series, World Health Organization, no. 31 (2001)
14. L.R. Baden, H.M. El Sahly, B. Essink, et al., Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. New Engl. J. Med. **384**, 403–416 (2021)
15. C. Baraniuk, Receptors for SARS-CoV-2 present in wide variety of human cells (2020). TheScientist. https://www.the-scientist.com/news-opinion/receptors-for-sars-cov-2-present-in-wide-variety-of-human-cells-67496. Accessed 13 Apr 2021
16. N. Bellomo, A. Bellouquid, L. Gibelli, N. Outada, *A Quest Towards a Mathematical Theory of Living Systems* (Birkhäuser, New York, 2017)
17. N. Bellomo, R. Bingham, M. Chaplain, G. Dosi, G. Forni, D. Knopoff, J. Lowengrub, R. Twarock, M. Virgillito, A multiscale model of virus pandemic: heterogeneous interactive entities in a globally connected world. Math. Model Methods Appl. Sci. **30**, 4591–1691 (2020)
18. N. Bellomo, D. Burini, G. Dosi, L. Gibelli, D. Knopoff, N. Outada, P. Terna, M.E. Virgillito, What is life? A perspective of the mathematical kinetic theory of active particles. Math. Models Methods Appl. Sci. **30**, 1821–1866 (2021). https://doi.org/10.1142/S0218202521500408
19. N. Bellomo, D. Burini, N. Outada, Pandemics of mutating virus and society: a multi-scale active particles approach. approach.Phil. Trans. R. Soc. A **380**, 20210161 (2022)
20. L. Billings, L. Mier-y-Teran-Romero, B. Lindley, I. Schwartz, Intervention-based stochastic disease eradication (2013). arXiv:1303.5614v1
21. K. Blyuss, Y. Kyrychko, Effects of latency and age structure on the dynamics and containment of COVID-19. J. Theor. Biol. **513**, 110587 (2021)
22. A. Carloni, V. Poletti, L. Fermo, N. Bellomo, M. Chilosi, Heterogeneous distribution of mechanical stress in human lung: a mathematical approach to evaluate abnormal remodeling. J. Theor. Biol. **332**, 136–140 (2013)
23. M. Cecconi, G. Forni, A. Mantovani, COVID-19: an executive report April 2020 update. Accademia Nazionale dei Lincei, Commissione Salute (2020)
24. D. Cucinotta, M. Vanelli, WHO declares COVID-19 a pandemic. Acta Biomed. **91**, 157–160 (2020)
25. N. Dagan, N. Barda, E. Kepten, et al., BNT162b2 mRNA Covid-19 vaccine in a nationwide mass vaccination setting. New Engl. J. Med. **384**, 1412–1423 (2021)
26. N.G. Davies, P. Klepac, Y. Liu, K. Prem, M. Jit, Age-dependent effects in the transmission and control of COVID-19 epidemics. Nat. Med. **26**, 1205–1211 (2020)
27. N.G. Davies, A.J. Kucharski, R.M. Eggo, A. Gimma, W.J. Edmunds, Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. Lancet Publ. Health **5**, e375–e385 (2020)
28. European Medicines Agency, COVID-19 vaccines: authorised. https://www.ema.europa.eu/en/human-regulatory/overview/public-health-threats/coronavirus-disease-covid-19/treatments-vaccines/vaccines-covid-19/covid-19-vaccines-authorised. Accessed 12 May 2021
29. H. Gang, Stationary solution of master equations in the large-system-size limit. Phys. Rev. A **36**, 5782–5790 (1987)
30. D.T. Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. J. Comput. Phys. **22**, 403–434 (1976)
31. D.T. Gillespie, Monte Carlo simulation of random walks with residence time dependent transition probability rates. J. Comput. Phys. **28**, 395–407 (1978)
32. E. Goldstein, M. Lipsitch, M. Cevik, On the effect of age on the transmission of SARS-CoV-2 in households, schools and the community. J. Infect. Dis. **223**, 362–369 (2020)
33. R. Heiland, Y. Wang, P. Macklin, Prototype 2-D multicellular simulation of COVID19 (2021). https://nanohub.org/resources/32987/
34. Johns Hopkins Coronavirus Resource Center, COVID-19. Johns Hopkins University & Medicine. https://coronavirus.jhu.edu/us-map. Accessed 15 Apr, 2020

35. L. Mateus, P. Ghaffari, U. Skwara, F. Rocha, M. Aguiar, D. Masoero, N. Stollenwerk, Semiclassical approximations of stochastic epidemiological processes towards parameter estimation using as prime example the SIS system with import. Ecol. Complex. **27**, 63–73 (2016)
36. L. Mateus, D. Masoero, F. Rocha, M. Aguiar, U. Skwara, P. Ghaffari, J.C. Zambrini, N. Stollenwerk, Epidemiological models in semiclassical approximation: an analytically solvable model as a test case. Math. Methods Appl. Sci. **39**, 4914–4922 (2016)
37. E. McBryde, J. Trauer, A. Adekunle, R. Ragonnet, M. Meehan, Stepping out of lockdown should start with school re-openings while maintaining distancing measures. Insights from mixing matrices and mathematical models (2020). https://medRxiv:2020.05.12.20099036
38. P. Musiani, G. Forni, Basic immunology 2020, in ISSUU (2020). https://issuu.com/guidoforni5/docs/2019o
39. M.A. Nowak, R. May, *Virus Dynamics: Mathematical Principles of Immunology and Virology* (Oxford University Press, Oxford, 2001)
40. A. Pinto, M. Aguiar, J. Martins, N. Stollenwerk, Dynamics of epidemiological models. Acta Biotheor. **58**(4), 381–389 (2010)
41. F. Pollack, S. Thomas, N. Kitchin, et al., Safety and efficacy of the BTN162b2 mRNA Covid-19 vaccine. New Engl. J. Med. **383**(27), 2603–2615 (2021). https://doi.org/10.1056/NEJMoa2034577
42. K. Prem, A.R. Cook, M. Jit, Projecting social contact matrices in 152 countries using contact surveys and demographic data. PLoS Comput. Biol. **13**, 1–21 (2017)
43. M. Roser, H. Ritchie, E. Ortiz-Ospina, J. Hasell, Coronavirus pandemic (COVID-19) (2021). OurWorldInData.org. https://ourworldindata.org/coronavirus
44. G. Rozhnova, C.H. van Dorp, P. Bruijning-Verhagen, et al., Model-based evaluation of school- and non-school-related measures to control the COVID-19 pandemic. Nat. Commun. **12**, 1614 (2021)
45. J. Sadoff, G. Gray, A. Vandebosch, et al., Safety and efficacy of single-dose Ad26.COV2.S vaccine against Covid-19. New Engl. J. Med. **384**, 2187 (2021). https://doi.org/10.1056/NEJMoa2101544
46. J. Shang, Y. Wan, C. Luo, G. Ye, Q. Geng, A. Auerbach, F. Li, Cell entry mechanisms of SARS-CoV-2. P. Natl. Acad. Sci. USA **117**(21), 11727–11734 (2020)
47. N. Stollenwerk, K.M. Briggs, Master equation solution of a plant disease model. Phys. Lett. A **274**, 84–91 (2000)
48. N. Stollenwerk, V. Jansen, *Population Biology and Criticality: From Critical Birth–Death Processes to Self-Organized Criticality in Mutation Pathogen Systems* (World Scientific, London, 2011)
49. N. Stollenwerk, S. van Noort, J. Martins, M. Aguiar, F. Hilker, A. Pinto, G. Gomes, A spatially stochastic epidemic model with partial immunization shows in mean field approximation the reinfection threshold. J. Biol. Dynam. **4**(6), 634–649 (2010)
50. N. Stollenwerk, P. Fuentes Sommer, B. Kooi, L. Mateus, P. Ghaffari, M. Aguiar, Hopf and torus bifurcations, torus destruction and chaos in population biology. Ecol. Complex. **30**, 91–99 (2017)
51. N. Stollenwerk, J. Bidaurrazaga Van-Dierdonck, J. Mar, I. Eguiguren Arrizabalaga, N. Cusimano, D. Knopoff, V. Anam, M. Aguiar, The interplay between subcritical fluctuations and import: understanding COVID-19 epidemiological dynamics, medRxiv (2021). https://doi.org/10.1101/2020.12.25.20248840
52. N. Stollenwerk, J. Mar, J. Bidaurrazaga Van-Dierdonck, O. Ibarrondo, C. Estadilla, M. Aguiar, Modeling COVID-19 vaccine efficacy and coverage towards herd-immunity in the Basque Country, Spain. medRxiv (2021). https://doi.org/10.1101/2021.07.12.21260390
53. The Epidemiological SHARUCD Model Dashboard, https://wp.bcamath.org/news/en/epidemiological-sharucd-model/#introduction. Accessed 23 Jul 2021
54. N.G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1992)

55. M. Voysey, S. Clemens, S. Mahdi, et al., Safety and efficacy of the ChAdOx nCov-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomized controlled trials in Brazil, South Africa and the UK. Lancet **397**, 10269 99–111 (2021)
56. Y. Wang, et al., Rapid community-driven development of a SARS-CoV-2 tissue simulator. bioRxiv (2020). https://doi.org/2020.04.02.019075
57. WHO Coronavirus (COVID-19) Dashboard. https://covid19.who.int/ Accessed 12 May 2021
58. World Bank, COVID-19 to Plunge global economy into worst recession since world war II. https://www.worldbank.org/en/news/press-release/2020/06/08/covid-19-to-plunge-global-economy-into-worst-recession-since-world-war-ii. Accessed 12 May 2021
59. World Health Organization, Naming the coronavirus disease (COVID-19) and the virus that causes it. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/technical-guidance/naming-the-coronavirus-disease-%28covid-2019%29-and-the-virus-that-causes-it Accessed 31 Jan 2021
60. World Health Organization, WHO announces COVID-19 outbreak a pandemic. https://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/news/news/2020/3/who-announces-covid-19-outbreak-a-pandemic Accessed 31 Jan 2021
61. World Health Organization, Coronavirus disease (COVID-19) weekly epidemiological update and weekly operational update https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports. Accessed 23 July 2021

Kinetic Modelling of Epidemic Dynamics: Social Contacts, Control with Uncertain Data, and Multiscale Spatial Dynamics

Giacomo Albi, Giulia Bertaglia, Walter Boscheri, Giacomo Dimarco, Lorenzo Pareschi, Giuseppe Toscani, and Mattia Zanella

1 Introduction

The recent COVID-19 pandemic has brought mathematical models in epidemiology to unprecedented scientific exposure. Several research groups in all parts of the planet have ventured into the construction and use of mathematical models capable of correctly describing the progress of the epidemic. Such predictive mathematical models are considered critical to understanding the course of the epidemic and planning effective control strategies. Most of the proposed models stem from the compartmental approach originally proposed by Kermack and McKendrick [65, 75], i.e., the classic SIR model based on partitioning the population into susceptible, infected, and recovered.

In compartmental models, the population is divided into groups, where each group or compartment represents a stage of progression in the individual's disease or health. The resulting mathematical model is typically based on a deterministic system of ordinary differential equations (ODEs) that characterizes the transition rates from one compartment to another. The choice of which compartments to

G. Albi
Department of Computer Science, University of Verona, Verona, Italy
e-mail: giacomo.albi@univr.it

G. Bertaglia · W. Boscheri · G. Dimarco · L. Pareschi (✉)
Department of Mathematics and Computer Science, University of Ferrara, Ferrara, Italy

Center for Modelling Computing and Simulation, University of Ferrara, Ferrara, Italy
e-mail: giulia.bertaglia@unife.it; walter.boscheri@unife.it; giacomo.dimarco@unife.it; lorenzo.pareschi@unife.it

G. Toscani · M. Zanella
Department of Mathematics, University of Pavia, Pavia, Italy
e-mail: giuseppe.toscani@unipv.it; mattia.zanella@unipv.it

N. Bellomo, M. A. J. Chaplain (eds.), *Predicting Pandemics in a Globally Connected World, Volume 1*, Modeling and Simulation in Science, Engineering and Technology, https://doi.org/10.1007/978-3-030-96562-4_3

include in a model depends on the characteristics of the particular disease being modelled and the purpose of the model. Recently, numerous extensions of the SIR compartmentalization have been proposed to deal with the specificity of SARS-CoV-2 infection. Such generalizations involve additional compartments to better fit the available experimental data in order to improve the description of disease progression and epidemic characteristics [28, 59, 74, 82, 96, 98, 112].

Most of these models describe the temporal evolution of the epidemic spread only in terms of the average numerical density of individuals in each compartment, thus neglecting other relevant structural properties of individuals, such as their age, wealth status, social contacts, and spatial movements, in favor of an assumption of population and territorial homogeneity. Structured epidemic models have been considered by various authors, especially in connection with age dependence, and are a classical topic in mathematical epidemiology [65, 66]. The evolution of these structural quantities, however, is seldom accounted, except for birth and death rates in the context of age-dependent models.

On the other hand, kinetic models characterized by systems of partial differential equations (PDEs), recently introduced and studied in the social sciences, have shown the ability to accurately describe complex social phenomena such as opinion formation among individuals, the creation of wealth distributions, and the emergence of contacts on social networks [6, 37, 56, 58, 63]. See, for instance, the recent monographs and collections [91, 95]. Similarly, the use of kinetic theory has proven to be very useful in designing feedback controlled models in various fields of social sciences [2, 4, 48] and in modelling the movement dynamics of individuals at different scales [7, 11, 14, 34, 108].

In addition, available experimental data are often affected by large uncertainty, which must therefore be considered as part of the process of modelling the infectious disease and simulating the potential epidemic scenarios and control strategies [29, 33, 106]. A large amount of research in this direction has been recently carried out in the field of hyperbolic and kinetic equations and it is therefore natural to rely on this scientific background to design new models and numerical methods able to deal efficiently with the presence of uncertain data [21, 68, 71, 94, 101, 122]. We also mention some other related research based on modelling the diffusion of COVID-19 using PDEs. Specifically, the multiscale approach in [15], the age- and space-structured model in [35], and the space-dependent models in [62, 107, 116, 117].

In this survey, we will address these issues in light of a number of recent results in the area of epidemiological modelling using kinetic equations with a focus on applications to the COVID-19 pandemic. More precisely, our presentation is organized into three parts according to the topics covered. In the first part we will focus on the interplay between the evolution of the pandemic and the presence of a social characteristic capable of significantly influencing its behavior and/or that is itself influenced by the pandemic. In the specific case we will consider the influence

of the distribution of contacts [45, 124, 125] and the distribution of wealth among individuals [44].

Next, in the second part our attention will shift to the importance of possible control actions aimed at containing the pandemic [10, 74]. In this case it is of fundamental importance to take into account in the modelling phase any uncertain data that can significantly change the epidemic scenarios. A particularly significant problem we will address is that of control actions through containment measures based on different age-dependent social interaction functions, e.g., at home, at work, or at leisure. Results for various countries affected by the pandemic will illustrate the effectiveness of the proposed methodology [8, 9, 46].

The last part of this survey is devoted to the challenging case of evaluating the impact of an infectious disease at a spatial level, including small-scale urban dynamics and large-scale regional dynamics. The inclusion of the spatial component in epidemiological systems is indeed crucial especially when there is a need to consider spatially heterogeneous interventions, as was and still is the case for the control of the spread of COVID-19 [39, 99, 105]. These problems will be addressed both in the case of dynamics on networks connecting different cities [18], and in the case of completely two-dimensional dynamics at regional level [23]. Applications of these models to the early stages of the COVID-19 pandemic in Italy will also be illustrated [19, 20]. Some open issues and future developments are also discussed at the end of this review along with detailed references to the data sources used in the simulations.

2 Kinetic Modelling of Social Heterogeneity in Epidemic Dynamics

We discuss in this first part an enhancement of the classical compartmental description of epidemic spread that takes into account statistical aspects of the social behavior of individuals [44, 45, 124, 125]. The approach has its roots in the kinetic theory of socio-economic modelling through interacting agents (see the monograph [95] for further details). This permits to correlate the social behavior of agents with the dynamics of infection.

For simplicity we will develop our arguments for the simple SIR model where the entire population is divided into three classes: susceptible (S), infected (I), and recovered (R) individuals. It should be noted, however, that the ideas developed in this section can be extended to more complex compartmental epidemic models like the ones considered in [25, 28, 40, 57, 59]. See also Sects. 3 and 4 of the present survey for some generalizations to more realistic compartmental models including the effects of asymptomatic individuals.

Under a homogeneous mixing assumption, the time evolution of the SIR model reads

$$\begin{aligned} \frac{dS(t)}{dt} &= -\beta S(t)I(t), \\ \frac{dI(t)}{dt} &= \beta S(t)I(t) - \gamma I(t), \\ \frac{dR(t)}{dt} &= \gamma I(t), \end{aligned} \tag{1}$$

where β is the average number of contacts per person per time, multiplied by the probability of disease transmission in a contact between a susceptible and an infectious person, and γ is the transition rate of infected to the recovered compartment. In this situation, it can be shown that the dynamics of the infectious class depends on the ratio $R_0 = \beta/\gamma$, the so-called basic reproduction number. In this case, it is known that $I(t) \to 0$, while $S(t) \to S^\infty \in [0, \gamma/\beta]$ solution of

$$I(0) + S(0) - S^\infty + \frac{\gamma}{\beta} \log\left(\frac{S^\infty}{S(0)}\right) = 0.$$

We refer to [65] for an introduction on compartmental modelling in epidemiology.

The heterogeneity of the social structure, which impacts the diffusion of the infective disease, is characterized by the variable $w \in \mathbb{R}_+$, characterizing its social state and whose components summarize, for example, the age of the individual, its number of social connections, or its economic status [64, 65]. For a large system of interacting individuals in a structured population, its statistical description is obtained through the introduction of the distribution functions $f_J(w,t)$, $t \geq 0$, denoting the probability of having an individual with the social characteristic w in the class J, where $J \in \{S, I, R\}$ and such that

$$f_S(w,t) + f_I(w,t) + f_R(w,t) = f(w,t), \qquad \int_{\mathbb{R}_+} f(w,t)dw = 1. \tag{2}$$

As a consequence, the quantities

$$S(t) = \int_{\mathbb{R}^+} f_S(w,t)\,dw, \quad I(t) = \int_{\mathbb{R}^+} f_I(w,t)\,dw, \quad R(t) = \int_{\mathbb{R}^+} f_R(w,t)\,dw, \tag{3}$$

denote the fractions of susceptible, infected, and recovered subjects. In the above setting, the time evolution of the functions $f_J(w,t)$, $J \in \{S, I, R\}$, is obtained by supplementing the epidemiological partitioning in (1) with the dynamics originating the formation of social heterogeneity by local interactions. Following [44, 95] this merging results in the system

$$\begin{aligned}
\frac{\partial f_S(w,t)}{\partial t} &= -K(f_S, f_I)(w,t) + Q_S(f_S; f_I; f_R)(w,t),\\
\frac{\partial f_I(w,t)}{\partial t} &= K(f_S, f_I)(w,t) - \gamma(w) f_I(w,t) + Q_I(f_S; f_I; f_R)(w,t),\\
\frac{\partial f_R(w,t)}{\partial t} &= \gamma(w) f_I(w,t) + Q_R(f_S; f_I; f_R)(w,t),
\end{aligned} \tag{4}$$

where

$$K(f_S, f_I)(w,t) = f_S(w,t) \int_{\mathbb{R}_+} \beta(v,w) f_I(v,t) dv \tag{5}$$

represents a nonlinear incidence rate characterized by the number of contacts $\beta(v,w)$ between an infectious individual with social characteristic v and a susceptible individual with social characteristic w, while $\gamma(w)$ defines the transition to the recovered compartment of infectious individuals with social feature w.

Note that, when the epidemic parameters are independent of the social feature, i.e., $\beta(\cdot,\cdot) = \beta$ and $\gamma(\cdot) = \gamma$, thanks to conservation of the number of individuals in each compartment during the evolution of social traits, by direct integration of (4) against w the mass densities (3) satisfy the classical SIR model (1).

The choice in (5) also implies a different definition of the basic reproduction number. Due to the conservation of the total number of individuals in the operators $Q_J(f_S; f_I; f_R)(w,t)$, $J \in \{S, I, R\}$, we define a time-dependent effective reproduction number

$$R_0(t) = \frac{\int_{\mathbb{R}_+^2} \beta(v,w) f_S(w,t) f_I(v,t)\, dv\, dw}{\int_{\mathbb{R}_+} \gamma(w) f_I(w,t)\, dw}, \tag{6}$$

which characterizes the change in the number of infected at time t.

In (4) the operators $Q_J(f_S; f_I; f_R)$, J in $\{S, I, R\}$, describe the evolution of social traits by interactions among agents in the various compartments and the formation of the corresponding equilibrium distributions $f_J(w)^\infty$, $J \in \{S, I, R\}$, such that

$$Q_J(f_S^\infty; f_I^\infty; f_R^\infty)(w) = 0, \qquad J \in \{S, I, R\}. \tag{7}$$

The explicit computation of the equilibrium solutions of (7) is extremely difficult in general, as it depends strongly on the evolution dynamics of the specific social feature under consideration. Knowledge of such equilibrium solutions, however, is of paramount importance to gain some understanding of the dynamics and derive simplified reduced-order models. In the following, we will describe in more detail the case of social heterogeneity based on the formation of suitable contact distributions [45] and the impact of the epidemic on the wealth distribution of individuals [44].

2.1 Modelling Contact Heterogeneity

Let us first consider a kinetic system that suitably describes the spreading of an infectious disease under the dependence of the contagiousness parameters on the number of social contacts of the agents. Aiming to understand social contacts' effects on the dynamics, we will not consider in the sequel the role of other sources of possible heterogeneity in the disease parameters (such as the personal susceptibility to a given disease), which could be derived from the classical epidemiological models, suitably adjusted to account for new information [41, 93, 115]. Therefore, we denote by $f_S(w,t)$, $f_I(w,t)$, and $f_R(w,t)$, the distributions at time $t > 0$ of the number of social contacts of the population of susceptible, infected, and recovered individuals.

For a given constant $\alpha > 0$, we denote with $m_{J,\alpha}(t)$, $J \in \{S, I, R\}$, the local moments of order α for the distributions of the number of contacts in each class conveniently divided by the mass of the class

$$m_{J,\alpha}(t) = \frac{1}{J(t)} \int_{\mathbb{R}^+} w^\alpha f_J(w,t)\, dw, \quad J \in \{S, I, R\}. \tag{8}$$

Unambiguously, we will indicate the local mean values, corresponding to $\alpha = 1$, by $m_J(t)$, $J \in \{S, I, R\}$.

In what follows, we assume that the various classes in the model act differently in the social process constituting the contact dynamics. Specifically, we will consider $\gamma(w) \equiv \gamma > 0$ and the contact function $\beta(v,w)$ as a nonnegative increasing function with respect to the number of contacts v and w of infected and susceptible individuals, respectively. The choice

$$\beta(v,w) = \bar{\beta} v^\alpha w^\alpha \tag{9}$$

with constant $\alpha, \bar{\beta} > 0$ corresponds to considering an incidence rate dependent on the product of the number of social contacts.

2.1.1 Kinetic Model for Contact Formation

To define the dynamics of contacts, we can exploit the results of [43, 61, 95, 113] to obtain a mathematical formulation of the formation of social contacts. In full generality, we assume that individuals in different compartments can have a different mean number of contacts. Then, the microscopic updates of social contacts from w to w'_J of individuals in the class $J \in \{S, I, R\}$ will be taken of the form:

$$w'_J = w - \Phi^\varepsilon_\delta(w/m_J)w + \eta_\varepsilon w, \qquad J \in \{S, I, R\}, \tag{10}$$

where for compactness and simplicity of notation we used the subscript J on the different compartments and kept implicit the dependence on ε in w'_J.

In a single update (interaction), the number w of contacts can be modified for two reasons, expressed by two terms, both proportional to the value w. In the first one, the function $\Phi^{\varepsilon}_{\delta}(\cdot)$, which takes both positive and negative values, characterizes the typical and predictable variation of the social contacts of agents, namely the personal social behavior of agents. The quantity η_ε is a random variable of zero mean and bounded variance of order $\varepsilon > 0$, expressed by $\langle \eta_\varepsilon \rangle = 0$, $\langle \eta_\varepsilon^2 \rangle = \varepsilon\sigma^2$, where $\langle \cdot \rangle$ denotes the expectation. Furthermore, we assume that η_ε has finite moments up to order three.

The function $\Phi^{\varepsilon}_{\delta}$ plays the role of the *value function* in the prospect theory of Kahneman and Tversky [72, 73]. See also [37, 87, 88, 95] for a related use of the value function in the dynamics of investment propensity. The main hypothesis on which this function is built is that, in relationship with the mean value w_J, $J \in \{S, I, R\}$, it is considered normally easier to increase the value of w (individuals look for larger networks) than to decrease it (people maintain as much connections as possible). In terms of the variable $s = w/m_J$ we consider then as in [43] the class of value functions obeying to the above general rule given by:

$$\Phi^{\varepsilon}_{\delta}(s) = \lambda \frac{e^{\varepsilon(s^{\delta}-1)/\delta} - 1}{e^{\varepsilon(s^{\delta}-1)/\delta} + 1}, \quad s \geq 0, \tag{11}$$

where the value λ denotes the maximal amount of variation of w that agents will be able to obtain in a single interaction

$$-\lambda \leq \Phi^{\varepsilon}_{\delta}(s) \leq \lambda,$$

so that the choice $\lambda < 1$ implies that, in the absence of randomness, the value of w'_J remains positive if w is positive. In (11) the parameter $0 < \delta \leq 1$ is a suitable constant characterizing the intensity of the individual behavior, while $\varepsilon > 0$ is related to the intensity of the interaction. We observe that $\varepsilon \ll 1$ corresponds to small variations of the expected difference $\langle w'_J - w \rangle$.

Thus, for a given density $f_J(w,t)$, $J \in \{S, I, R\}$, the operators $Q_J(f_S; f_I; f_R)(w,t)$ on the right hand side of (4) have a linear structure, depending only on compartment J, characterized by the microscopic interaction (10). Denoting by $Q^{\varepsilon}_J(f_J) = Q_J(f_S; f_I; f_R)$, the interaction terms can be conveniently written in weak form by integration against a smooth function $\varphi(w)$ as [32, 95]

$$\int_{\mathbb{R}_+} \varphi(w) Q^{\varepsilon}_J(f_J)(w,t)\, dw = \Big\langle \int_{\mathbb{R}_+} B(w)\big(\varphi(w'_J) - \varphi(w)\big) f_J(w,t)\, dw \Big\rangle. \tag{12}$$

The above operators quantify the variation in density, at a given time, of individuals in the class $J \in \{S, I, R\}$ that modify their value from w to w'_J (r.h.s with negative sign) and agents that change their value from w'_J to w (r.h.s. with positive sign). Here, the expectation $\langle \cdot \rangle$ takes into account the presence of the random parameter

η_ε in the microscopic interaction (10), while the function $B(w)$ measures the interaction frequency. For example, the choice $B(w) = 1/w$, which will be used in the sequel, assigns a low probability to interactions where individuals already have a large number of contacts and assigns a high probability to contact transitions when the value of the variable w is small.

2.1.2 Quasi-Invariant Scaling and Steady States

Let us focus on the dynamics of social contacts alone in (4), namely by ignoring the epidemiological terms, and scale time as $t \to t/\varepsilon$, in accordance with the parameter ε that measures the intensity of changes in the number of contacts defined by (10). Thus, small values of ε correspond to the case in which elementary interactions (10) produce minimal modification of the number of social contacts and at the same time their frequency increases like $1/\varepsilon$. This scaling is usually referred to as quasi-invariant scaling in kinetic socio-economic modelling [95]. A general view about this asymptotic passage from kinetic equations based on general interactions toward Fokker–Planck-type equations can be found in [55].

Then, as a result of the scaling, the distribution f_J, $J \in \{S, I, R\}$ is solution of the following problem in weak form:

$$\begin{aligned} \frac{d}{dt}\int_{\mathbb{R}_+} \varphi(w) f_J(w,t) dw &= \frac{1}{\varepsilon}\int_{\mathbb{R}_+} \varphi(w) Q_J^\varepsilon(f_J)(w,t)\, dw \\ &= \frac{1}{\varepsilon}\Big\langle \int_{\mathbb{R}_+} B(w)\big(\varphi(w'_J) - \varphi(w)\big) f_J(w,t)\, dw \Big\rangle. \end{aligned} \tag{13}$$

Now, let us concentrate on the analysis of the asymptotic states of the social contact dynamics when $\varepsilon \to 0$. To this aim, note that, from the definition of Φ_δ^ε in (11) and the assumptions on the noise term η_ε we have

$$\lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\Phi_\delta^\varepsilon\left(\frac{w}{m_J}\right) = \frac{\lambda}{2\delta}\left[\left(\frac{w}{m_J}\right)^\delta - 1\right], \qquad \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\langle \eta_\varepsilon^2\rangle = \sigma^2. \tag{14}$$

We can Taylor expand $\phi(w'_J)$ in (13) as

$$\varphi(w'_J) - \varphi(w) = (w'_J - w)\varphi'(w) + \frac{1}{2}(w'_J - w)^2\varphi''(w) + \frac{1}{6}(w'_J - w)^3\varphi'''(\hat{w}_J),$$

with $\hat{w}_J$ a suitable value between w'_J and w_j. Hence, inserting the above expansion in (13) and using the microscopic relation (10), as $\varepsilon \to 0$ by standard arguments we can prove that the scaled dynamics (13) can be approximated by the corresponding Fokker–Planck formulation [45, 95]. More precisely, it can be shown that f_J, $J \in \{S, I, R\}$ converges to a solution of

$$\frac{d}{dt}\int_{\mathbb{R}_+}\varphi(w)f_J(w,t)dw =$$
$$\int_{\mathbb{R}_+}\left\{-\varphi'(w)\,\frac{\lambda\,w^{1-\delta}}{2\delta}\left[\left(\frac{w}{m_J}\right)^{\delta}-1\right]+\frac{\sigma^2}{2}\varphi''(w)\,w^{2-\delta}\right\}f_J(w,t)\,dw.$$

Integrating back by parts, the limit equation in strong form coincides with the Fokker–Planck equation

$$\frac{\partial}{\partial t}f_J(w,t)=\tilde{Q}_J(f_J)(w,t),\qquad J\in\{S,I,R\},$$

where

$$\tilde{Q}_J(f_J)(w,t)=$$
$$\frac{\lambda}{2\delta}\frac{\partial}{\partial w}\left\{w^{1-\delta}\left[\left(\frac{w}{m_J}\right)^{\delta}-1\right]f_J(w,t)\right\}+\frac{\sigma^2}{2}\frac{\partial^2}{\partial w^2}(w^{2-\delta}f_J(w,t)),\tag{15}$$

complemented with no-flux boundary conditions at $w=0$

$$\frac{\partial}{\partial w}(w^{2-\delta}f_J(w,t))\Big|_{w=0}=0.\tag{16}$$

Following [45] we can compute the explicit equilibrium distribution of the Fokker–Planck model. Let us first observe that Eq. (15) preserves the total number of individuals and the average number of contacts' mean values m_J, $J\in\{S,I,R\}$, in each compartment. Thus, assuming that the mass of the initial distribution is one and by setting $\mu=\lambda/\sigma^2$, the equilibria can be expressed by the functions

$$f_J^{\infty}(w)=C_J(m_J,\delta,\mu)w^{\mu/\delta+\delta-2}\exp\left\{-\frac{\mu}{\delta^2}\left(\frac{w}{m_J}\right)^{\delta}\right\},\qquad J\in\{S,I,R\},\tag{17}$$

where $C_J>0$ is a normalization constant.

The Distribution of Contacts

A particular interesting case corresponds to the choice $\delta=1$ for which the steady states of unit mass are the Gamma densities

$$f_J^{\infty}(w;\theta,\mu)=\left(\frac{\mu}{m_J}\right)^{\mu}\frac{1}{\Gamma(\mu)}w^{\mu-1}\exp\left\{-\frac{\mu}{m_J}w\right\},\quad J\in\{S,I,R\}.\tag{18}$$

(continued)

With this particular choice, the mean values and the energies of the densities (18), $J \in \{S, I, R\}$, are given by:

$$\int_{\mathbb{R}^+} w\, f_J^\infty(w;\theta,\mu)\, dw = m_J, \qquad \int_{\mathbb{R}^+} w^2\, f_J^\infty(w;\theta,\mu)\, dw = \frac{\mu+1}{\mu} m_J^2. \tag{19}$$

It is important to note that the distribution (18) is in agreement with that observed experimentally in [16]. For this reason, in the rest of the section we will restrict to the case $\delta = 1$ (see [45] for a more in-depth discussion).

2.1.3 The Macroscopic Social-SIR Dynamics

Referring to Boltzmann's classical legacy concerning the fluid dynamic limits, using the knowledge of the equilibrium states of the kinetic model we can derive the corresponding macroscopic model [32]. The key assumption is that the dynamics leading to the contact formation is much faster than the epidemic dynamics. This corresponds to introduce the following scaling:

$$t \to t/\tau, \qquad \beta(v,w) \to \tau\beta(v,w), \qquad \gamma \to \tau\gamma,$$

being $\tau \ll 1$ the scaling parameter.

Hence, considering the linear Fokker–Planck operator (15) for $\delta = 1$ as a model for social interactions we can rewrite system (4) as follows:

$$\begin{aligned}
\frac{\partial f_S(w,t)}{\partial t} &= -K(f_S, f_I)(w,t) + \frac{1}{\tau}\tilde{Q}_S(f_S)(w,t),\\
\frac{\partial f_I(w,t)}{\partial t} &= K(f_S, f_I)(w,t) - \gamma f_I(w,\tau) + \frac{1}{\tau}\tilde{Q}_I(f_I)(w,t),\\
\frac{\partial f_R(w,t)}{\partial t} &= \gamma f_I(w,t) + \frac{1}{\tau}\tilde{Q}_R(f_R)(w,t).
\end{aligned} \tag{20}$$

The system (20) with no-flux boundary conditions at $w = 0$ contains all the information on the spreading of the epidemic in terms of the distribution of social contacts. Indeed, the knowledge of the densities $f_J(w,t)$, $J \in \{S, I, R\}$, allows to evaluate by integrations all moments of interest. Due to the incidence rate $K(f_S, f_I)$, as given in (5), the time evolution of the moments of the distribution functions is not explicitly computable, since the evolution of a moment of a certain order depends on the knowledge of higher order moments, thus producing a hierarchy of equations, like in classical kinetic theory of rarefied gases [32].

However, similarly to the derivation of the fluid dynamic limit we can assume the contact densities to be close to their equilibrium states (18).

Therefore, since for the choice in (9) we have

$$K(f_S, f_I)(w,t) = \bar{\beta} w \, f_S(w,t) m_I(t) \, I(\tau),$$

we can compute the time evolution of the number of individuals in each compartment, defined in (3), by integrating both sides of the equations in (20) with respect to w. Using the fact that the Fokker–Planck terms preserve the total number of individuals, we obtain the following system of macroscopic equations for the densities:

$$\begin{aligned} \frac{dS(t)}{dt} &= -\bar{\beta} \, m_S(t) m_I(t) S(t) I(t), \\ \frac{dI(t)}{dt} &= \bar{\beta} \, m_S(t) m_I(t) S(t) I(t) - \gamma I(t), \\ \frac{dR(t)}{dt} &= \gamma I(t). \end{aligned} \tag{21}$$

Next, taking the evolution of the first moment in (20) since the Fokker–Planck operators also preserve momentum, one obtains that the means $m_S(t)S(t)$, $m_I(t)I(t)$ in (21) satisfy the differential system

$$\begin{aligned} \frac{d}{dt}(m_S(t)S(t)) &= -\bar{\beta} \, m_{S,2}(t) m_I(t) S(t) I(t), \\ \frac{d}{dt}(m_I(t)I(t)) &= \bar{\beta} \, m_{S,2}(t) m_I(t) S(t) I(t) - \gamma m_I(t) I(t), \end{aligned} \tag{22}$$

which depends now on the second order moments.

The closure of system (21)–(22) can be obtained by resorting, at least formally, to the classical equilibrium assumption on the social interaction variable. Indeed, if $\tau \ll 1$ is sufficiently small, one can easily argue from the exponential convergence of the solution $f_J(w,t)$ of the Fokker–Planck equation toward the equilibrium $f_J^\infty(w;\theta,\mu)$, $J \in \{S, I, R\}$ (see [114] for example) that the solution remains sufficiently close to the corresponding Gamma density (18) for all times.

The equilibrium distribution $f_J^\infty(w;\theta,\mu)$ can then be inserted into system (22) and, recalling that for Gamma densities

$$m_{J,2}(t) = \frac{\mu+1}{\mu} m_J^2(t), \qquad J \in \{S, I, R\},$$

we can derive a closed system that governs the evolution of the local mean values

$$\begin{aligned}
\frac{dm_S(t)}{dt} &= -\frac{\bar{\beta}}{\mu} m_S(t)^2 m_I(t) I(t), \\
\frac{dm_I(t)}{dt} &= \bar{\beta} m_S(t) m_I(t) \left(\frac{1+\mu}{\mu} m_S(t) - m_I(t) \right) S(t), \\
\frac{dm_R(t)}{dt} &= \gamma \frac{I(t)}{R(t)} \left(m_I(t) - m_R(t) \right).
\end{aligned} \tag{23}$$

Therefore, the closure of the kinetic system (20) around a Gamma-type equilibrium of social contacts leads then to the system of six Eqs. (21)–(23) for the pairs of mass fractions $J(t)$ and local mean values $m_J(t)$, $J \in \{S, I, R\}$.

It is interesting to remark that system (23) is explicitly dependent on the positive parameter $\mu = \lambda/\sigma^2$, which measures the heterogeneity of the population in terms of the variance of the statistical distribution of social contacts. More precisely, small values of the constant μ correspond to high values of the variance, and thus to a larger heterogeneity of the individuals with respect to social contacts. This is an important point that is widely present and studied in the epidemiological literature [10, 27, 40, 41].

Absence of Heterogeneity
A limiting case of system (23) is obtained by letting the parameter $\mu \to +\infty$, which corresponds to pushing the variance to zero (absence of heterogeneity). In this case, if the whole population starts with a common number of daily contacts, say $\bar{w}$, it is immediate to show that the number of contacts remains fixed in time, thus reducing system (21) to a classical SIR model with contact rate $\bar{\beta}\bar{w}^2$. Hence this classical epidemiological model is contained in (21)–(23) and corresponds to consider the case of a population that, regardless of the presence of the epidemic, maintains the same fixed number of daily contacts.

2.1.4 A Social-SIR Model with Saturated Incidence Rate

We consider the case where the average number of social contacts of infected m_I is frozen to $\tilde{m}_I$ as an effect, for instance, of external interventions aimed at controlling the pandemic spread. In this case, for any $\alpha \geq 1$, one can explicitly solve the equation for the evolution of average contacts of susceptible individual as

$$\frac{d}{dt} m_S(t) = -\frac{\bar{\beta} c_\alpha}{\mu} m_S^{\alpha+1} \tilde{m}_I^\alpha I(t), \quad m_S(t) = \frac{m_S(0)}{\left(1 + \dfrac{c_\alpha \bar{\beta} \alpha m_S^\alpha(0)}{\mu} \tilde{m}_I^\alpha \displaystyle\int_0^t I(s) ds \right)^{1/\alpha}},$$

where $c_\alpha > 0$ is such that

$$\int_0^{+\infty} w^\alpha f_S^\infty(w)dw = c_\alpha S(t) w_S^\alpha.$$

Therefore, approximating the integral $\int_0^t I(s)\,ds \approx tI(t)$ we obtain the closed system for the evolution of mass fractions of the following type:

$$\begin{aligned}
\frac{dS(t)}{dt} &= -\tilde{\beta}\, S(t)I(t)H(I(t),t),\\
\frac{dI(t)}{dt} &= \tilde{\beta}\, H(I(t),t)S(t)I(t) - \gamma I(t),\\
\frac{dR(t)}{dt} &= \gamma I(t),
\end{aligned} \tag{24}$$

with $\tilde{\beta} = \bar{\beta} m_S(0)$ and which incorporates the generalized macroscopic incidence function

$$H(I(t),t) = \frac{1}{(1+\phi(t)I(t))^{1/\alpha}}, \tag{25}$$

with $\phi(t) = c_\alpha \alpha \bar{\beta} m_S^\alpha(0) t/\mu > 0$. The system (24) corresponds to models with saturated incidence rate, see [30]. We point the interested reader to [45] for a detailed discussion. See also Sect. 3 of this survey for a derivation of the saturated incidence function (25) as a feedback control functional.

2.1.5 Extrapolation of the Shape of the Incidence Rate from Data

In this section, we use the previous model to describe the COVID-19 pandemic in three different European countries: France, Italy, and Spain. The data we employ, concerning the actual number of infected, recovered, and deaths of COVID-19, are publicly available from the John Hopkins University GitHub repository. For the specific case of Italy, we considered instead the GitHub repository of the Italian Civil Protection Department (see Data Sources in Sect. 5.1). We adopted the fitting procedure described in [8, 45] that is based on a strategy with two optimization horizons (pre-lockdown and lockdown time spans) depending on the different strategies enacted by the governments of the considered European countries. Once the relevant epidemiological parameters have been estimated in the pre-lockdown time span, we successively proceeded with the estimation of the shape of the function H from the data.

We seek to understand numerically the dependencies of the function H on the number of infected. In particular, we consider the candidate incidence functions H_1, H_2, and H_3 defined as

$$H_1(I(t),t) = \frac{c}{1+\phi I(t)}, \quad H_2(I(t),t) = \frac{c}{1+\phi \int_0^t I(s)\,ds},$$

and

$$H_3(I(t),t) = \frac{c}{\left(1+\phi \int_0^t I(s)\,ds\right)^{1/\alpha}},$$

with $c > 0$, accordingly with (25) where ϕ and α are free parameters, which are determined through a least square minimization approach that best fits the estimated curve with conditions $\phi > 0, \alpha \geq 1$. The results of this procedure are presented in Fig. 1. We point the interested reader to [45] for a detailed discussion on the estimated parameters.

We can observe that the optimization gives acceptable results for the different forms of the incidence function especially in the right column of Fig. 1, where the functions H_2 and H_3 are clearly able to better explain the estimated values of H especially after the epidemic peak. Note that the fits of the model with the available data when H_3 is used are particularly good. This fact may indicate that people are rather fast to apply social distancing, and therefore to reduce their average number of contacts, whereas they tend to restore the pre-pandemic average contact rate more slowly, possibly due to further psychological effects.

2.2 *The Interplay Between Economy and the Pandemic*

The rapid spreading of the COVID-19 epidemic in western countries and the consequent lockdown measures assumed by the governments to control and limit its effects will unequivocally lead to important consequences for their economies. In this section, following [44], we introduce a mathematical framework to study the economic impact of the pandemic by integrating epidemiological dynamics with a kinetic model of wealth exchange.

The description of the evolution of the personal wealth in terms of kinetic-type equations revealed to be successful in the description of emerging wealth distributions; see [24, 36, 48, 95]. Clearly, an accurate quantification of the implications due to the pandemic in the distribution of wealth is an extremely difficult problem that requires knowledge of a large number of unknown variables and relationships between them. In an attempt to better understand the mechanisms underlying these dynamics we will consider simplified models that, while based on a few obvious universal characteristics, can be analyzed to provide answers about possible scenarios.

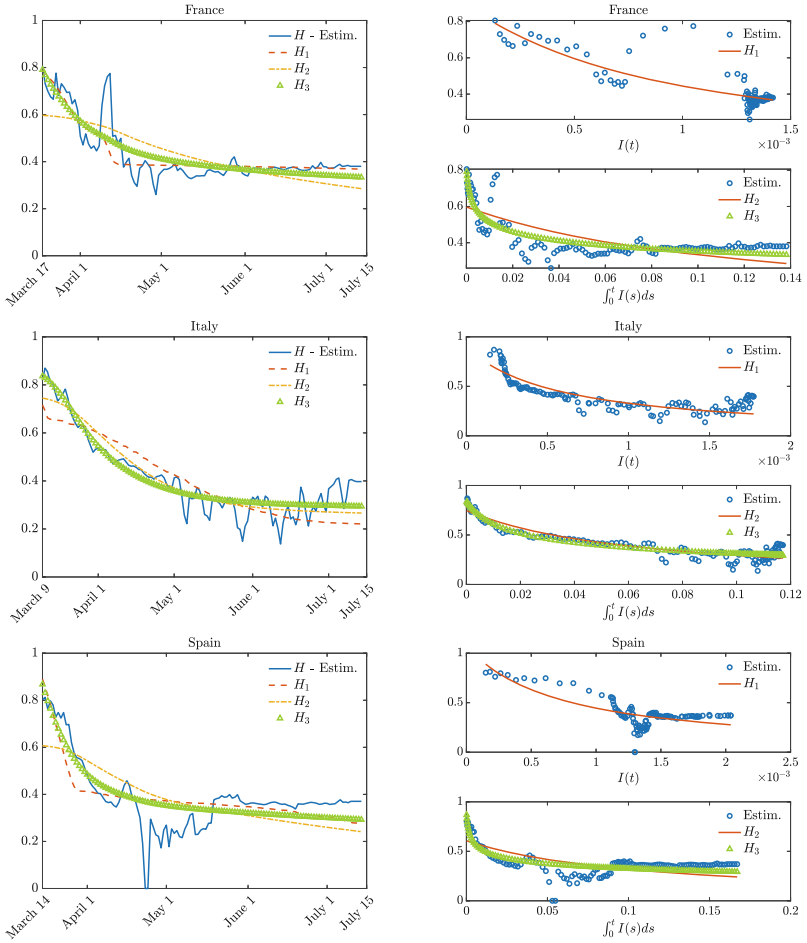


Fig. 1 Estimated shape of the function H in several European countries (left plots) and its dependency on the variables $I(t)$ and $\int_0^t I(s)ds$ (right plots)

2.2.1 Wealth Exchanges in Epidemic Modelling

The model considered in [44] has the same structure of the kinetic SIR-type model defined in (4) where now the state of an individual in each class at any instant of time $t \geq 0$ is completely characterized by the amount of wealth $w \in \mathbb{R}_+$. Therefore, $f_S(w,t)$, $f_I(w,t)$, and $f_R(w,t)$ are the distributions at time $t > 0$ of the amount of wealth of the population of susceptible, infected, and recovered individuals, respectively. The distribution of wealth of the whole population is then recovered as in (2) and consequently is assumed to be a probability density for all times $t \geq 0$.

Coherently with the previous notations we denote the relative mean wealths in each compartment as

$$m_J(t) = \frac{1}{J(t)} \int_{\mathbb{R}^+} w f_J(w,t)\, dw, \quad J \in \{S, I, R\}, \tag{26}$$

and the total mean wealth as

$$m(t) = \int_{\mathbb{R}^+} w f(w,t)\, dw.$$

We emphasize that the above notations differ from those originally used in [44], this in order to make them homogeneous with the notations used in Sect. 2.1.

In Eq. (4) the choice of a wealth-dependent recovery rate can be motivated by considering that wealth can buy access to better hospitals and better treatments, thus ensuring a higher chance of recovery. Also, a wealth-dependent contact rate, in the form of a decreasing function of the difference $|w - w_*|$, can be introduced to express that individuals with different degrees of wealth live in different environments, and this limits contacts in the presence of a marked difference. The interaction operators on the r.h.s. of (4) characterize the wealth evolution due to trading between agents of the same class, or between agents of different classes, and are built according to the CPT model [36] with the following structure:

$$Q_J(f_S; f_I; f_R) = \sum_{H \in \{S,I,R\}} Q_{JH}(f_J, f_H)(w,t), \qquad J \in \{S, I, R\}, \tag{27}$$

where $Q_{JH}(f_J, f_H)$ describes the changes of wealth in compartment J due to binary interactions among agents in compartments J and H, with $J, H \in \{S, I, R\}$.

In detail, an interaction between two individuals in compartments J and H with wealth pair (w, w_*) leads to a wealth pair (w'_{JH}, w'_{HJ}) defined by relations

$$\begin{aligned} w'_{JH} &= (1 - \lambda_J) w + \lambda_H w_* + \eta_{JH} w \\ w'_{HJ} &= (1 - \lambda_H) w_* + \lambda_J w + \eta_{HJ} w_*, \end{aligned} \qquad J, H \in \{S, I, R\} \tag{28}$$

where $\lambda_J, \lambda_H \in (0, 1)$ are transaction coefficients, while the market risk variables $\eta_{JH} \geq -\lambda_J$ and $\eta_{HJ} \geq -\lambda_H$ are independent and identically distributed random variables with zero mean and the same time-dependent variance $\sigma^2(t)$ (since we assume that the risk in the market does not depend on the particular class of trading agents).

The trade between agents has been modelled to include the idea that wealth changes hands for a specific reason: one agent intends to invest his wealth in some asset, property etc. in possession of his trade partner. Typically, such investments bear some speculative risk, and either provide the buyer with some additional wealth, or lead to the loss of wealth in a non-deterministic way. Relations (28)

couple the saving propensity parameter with some risky investment that yields an immediate gain or loss proportional to the current wealth of the investing agent. Hence $0 < \lambda_J < 1$, $J \in \{S, I, R\}$, are the parameters that identify the saving propensities $1 - \lambda_J$, namely the intuitive behavior that prevents the agents to put in a single trade the whole amount of his money. The choice $\lambda_R > \lambda_S$, for example, reflects the fact that susceptible individuals can be more cautious in the market and tend to save their wealth, since they understand that consuming and working less reduces the probability of infection [50]. On the other hand, infectious individuals have limited possibilities to act on the market and, as we will see, asymptotically disappear from the wealth dynamics. The time dependence of σ has been postulated by assuming that, in the presence of a significant spread of the epidemic, the risk variance tends to increase. This is in agreement, for example, to the market reactions we observed during the COVID-19 spreading at the announcements of the new numbers of infectious people in the various countries [127].

As already observed a convenient way to express the operators $Q_{JH}(f_J, f_H)$ is based on its weak form, namely the way the operator acts on observables [36, 95]. Let $\phi(v)$ denote a test function and let us define with $\langle\cdot\rangle$ the expected value with respect to the pair η_{JH}, η_{HJ} in the interaction process. Thus, for $J, H \in \{S, I, R\}$ we have

$$\int_{\mathbb{R}_+} \phi(w) Q_{JH}(f_J, f_H)(w,t)\, dw = \left\langle \int_{\mathbb{R}_+^2} (\phi(w'_{JH}) - \phi(w)) f_J(w,t) f_H(w_*,t)\, dw_*\, dw \right\rangle, \tag{29}$$

where w'_{JH} is defined by (28).

2.2.2 Fokker–Planck Scaling and Steady States

To analyze the asymptotic behavior of the model, it is useful to resort to the so-called quasi-invariant trading limit that permits to derive the corresponding Fokker–Planck description of the Boltzmann operators (29). To this aim, in a similar fashion to Sect. 2.1.2, following [36, 44, 55, 95], we scale the binary trades according to

$$\lambda_J \to \varepsilon \lambda_J, \quad J \in \{S, I, R\}, \quad \sigma \to \sqrt{\varepsilon}\sigma, \tag{30}$$

and similarly the functions governing the spread of the disease

$$\beta(w, w_*) \to \varepsilon \beta(w, w_*), \quad \gamma(w) \to \varepsilon \gamma(w), \tag{31}$$

and denote with $Q^{\varepsilon}_{JH}(\cdot,\cdot)$, $J, H \in \{S, I, R\}$, the scaled interaction terms.

The limit procedure induced by the above scaling corresponds to the situation in which are prevalent the exchanges of wealth that produce an extremely small

modification of wealths, but we are waiting enough time to still see the effects. In fact, rescaling time as $t \to t/\varepsilon$, for small values of ε, the Boltzmann-type operators converge to Fokker–Planck operator with variable coefficient of diffusion and linear drift. More precisely we have that for small values of $\varepsilon \ll 1$ (see [44] for details)

$$\frac{1}{\varepsilon} \sum_{H \in \{S,I,R\}} \int_{\mathbb{R}_+} Q^{\varepsilon}_{JH}(f_J, f_H)(w,t)\phi(w)\, dw \approx \int_{\mathbb{R}_+} \tilde{Q}_J(f_J)(w,t)\phi(w)\, dw,$$

where

$$\tilde{Q}_J(f_J)(w,t) = \frac{\partial}{\partial w}\left[\frac{\sigma(t)^2}{2}\frac{\partial}{\partial w}(w^2 f_J(w,t)) + (w\lambda_J - \bar{m}(t))\, f_J(w,t)\right], \tag{32}$$

with

$$\bar{m}(t) = \lambda_S m_S(t) S(t) + \lambda_I m_I(t) I(t) + \lambda_R m_R(t) R(t). \tag{33}$$

This gives the system

$$\begin{aligned}
\frac{\partial f_S(w,t)}{\partial t} &= -K(f_S, f_I)(w,t) + \tilde{Q}_S(f_S)(w,t),\\
\frac{\partial f_I(w,t)}{\partial t} &= K(f_S, f_I)(w,t) - \gamma f_I(w,t) + \tilde{Q}_I(f_I)(w,t),\\
\frac{\partial f_R(w,t)}{\partial t} &= \gamma f_I(w,t) + \tilde{Q}_R(f_R)(w,t).
\end{aligned} \tag{34}$$

It is immediate to verify that the above Fokker–Planck-type operators are mass and momentum preserving. Similarly to Sect. 2.1.3, one can analyze the equilibrium densities associated to the differential system

$$\frac{\sigma^2}{2}\frac{\partial (w^2 f_J^{\infty}(w))}{\partial w} + (w\lambda_J - \bar{m})\, f_J^{\infty}(w) = 0, \quad J \in \{S, I, R\},$$

to derive reduced order models for the evolution of the densities of susceptible, infectious and recovered individuals. As we will see in the next section, these equilibrium states have the shape of inverse Gamma distributions [78, 109].

2.2.3 The Formation of Bimodal Wealth Distributions

We verify in a simplified case that the Fokker–Planck system (34) possesses as stationary solutions inverse Gamma distributions that may generate a bimodal form of wealth distribution. Bimodal shapes are typical of situations of high stress in economy and are investigated starting from the Argentinian crisis of the first year

of the new century [58, 63]. This example also shows that a similar behavior can be expected in reason of the epidemic spreading.

Suppose that $\beta(w, w_*) = \beta$, $\gamma(w) = \gamma$, and $\sigma(t) = \sigma$ are constant. Then, integrating with respect to the wealth variable, thanks to conservation of the total wealth, we obtain that the relative mass densities satisfy the classical SIR model (1). In this case, it is known that $I(t) \to 0$, while $S(t) \to S^\infty \in [0, \gamma/\beta]$. Likewise, the system for the mean values reads

$$\frac{d(m_S(t)S(t))}{dt} = -\beta I(t) m_S(t)\, S(t) + (\bar{m}(t) - \lambda_S m_S(t))S(t), \tag{35}$$

$$\frac{d(m_I(t)I(t))}{dt} = \beta I(t) m_S(t)\, S(t) - \gamma m_I(t)\, I(t) + (\bar{m}(t) - \lambda_I m_I(t))I(t), \tag{36}$$

$$\frac{d(m_R(t)R(t))}{dt} = \gamma m_I(t)\, I(t) + (\bar{m}(t) - \lambda_R m_R(t))R(t). \tag{37}$$

Since, as $t \to +\infty$ we have $I(t) \to 0$, $m_S(t) \to m_S^\infty$, and $m_R(t) \to m_R^\infty$, the asymptotic values of the means satisfy

$$\lambda_R m_R^\infty = \lambda_S m_S^\infty,$$

together with the constraint $m_R^\infty\, R^\infty + m_S^\infty\, S^\infty = m$ by conservation of the total mean wealth. This gives the asymptotic values

$$m_S^\infty = \frac{\lambda_R}{\lambda_R S^\infty + \lambda_S R^\infty} m, \qquad m_R^\infty = \frac{\lambda_S}{\lambda_R S^\infty + \lambda_S R^\infty} m. \tag{38}$$

Thus, formally as $t \to \infty$ in the Fokker–Planck system (34) we get that the stationary states $f_S^\infty(w)$ and $f_R^\infty(w)$ are given by two inverse Gamma densities

$$f_S^\infty(w) = S^\infty \frac{\kappa^{\mu_S}}{\Gamma(\mu_S)} \frac{e^{-\frac{\kappa}{w}}}{w^{1+\mu_S}}, \qquad f_R^\infty(w) = R^\infty \frac{\kappa^{\mu_R}}{\Gamma(\mu_R)} \frac{e^{-\frac{\kappa}{w}}}{w^{1+\mu_R}} \tag{39}$$

with

$$\mu_S = 1 + 2\frac{\lambda_S}{\sigma^2}, \qquad \mu_R = 1 + 2\frac{\lambda_R}{\sigma^2}, \quad \kappa = (\mu_S - 1)m_S^\infty = (\mu_R - 1)m_R^\infty. \tag{40}$$

The details of the trading activity at the basis of the kinetic description allow to characterize the tails of the distributions from (40). Hence, a low value of the Pareto index is obtained in the presence of small values of the parameter λ_S, λ_R (small saving propensity of agents), or to high values of the parameter σ (highly risky market). Therefore, the asymptotic wealth distribution is the mixture of two inverse Gamma densities of mass S^∞ and R^∞, respectively,

$$f^\infty(w) = f_S^\infty(w) + f_R^\infty(w), \tag{41}$$

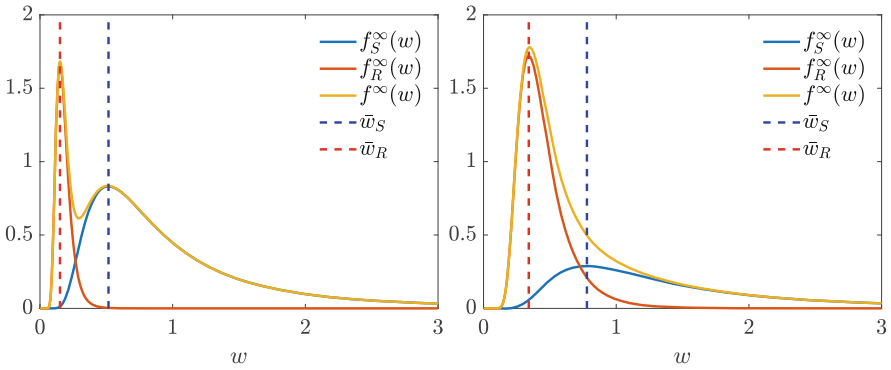


Fig. 2 Exact solutions for wealth distributions at the end of the epidemic (34) in the Fokker–Planck approximation for $\mu_S = 2.5$, $\mu_R = 7.0$, $S^\infty = 0.4$, and $R^\infty = 0.6$ (top) and $\mu_S = 2.5$, $\mu_R = 11.0$, $S^\infty = 0.8$, and $R\infty = 0.2$ (bottom)

with asymptotic means (38) and variances given by:

$$\mathrm{Var}_S^\infty = \frac{\kappa^2}{(\mu_S - 1)(\mu_S - 2)}, \qquad \mathrm{Var}_R^\infty = \frac{\kappa^2}{(\mu_R - 1)(\mu_S - 2)}, \qquad \mu_R, \mu_S > 2.$$

As a consequence, the wealth distribution has a bimodal structure, since the maximum of $f_S^\infty(w)$ and $f_R^\infty(w)$ are achieved, respectively, at the points

$$\bar{w}_S = \frac{\kappa}{\mu_S + 1}, \qquad \bar{w}_R = \frac{\kappa}{\mu_R + 1}. \tag{42}$$

We report in Fig. 2 the resulting profiles for various choices of $\mu_S < \mu_R$, and S^∞, R^∞. Note that the mixture of the two inverse Gamma densities (39) does not always result in an evident bimodal shape. Indeed, while the profile on the right of Fig. 2 is clearly bimodal, a different choice of parameters on the left produces a unimodal steady profile.

2.2.4 The Increase of Wealth Inequalities

Next, we compare the evolution of the wealth distribution of the system under more realistic hypotheses about the dependence of the risk coefficient σ on the epidemic spread. We consider the kinetic model (34) in the case of the following two infectious-dependent market risk coefficients:

$$\sigma_1(t) = \sigma_0(1 + \alpha I(t)), \qquad \sigma_2(t) = \sigma_0\left(1 + \alpha \int_0^t I(\tau)d\tau\right), \tag{43}$$

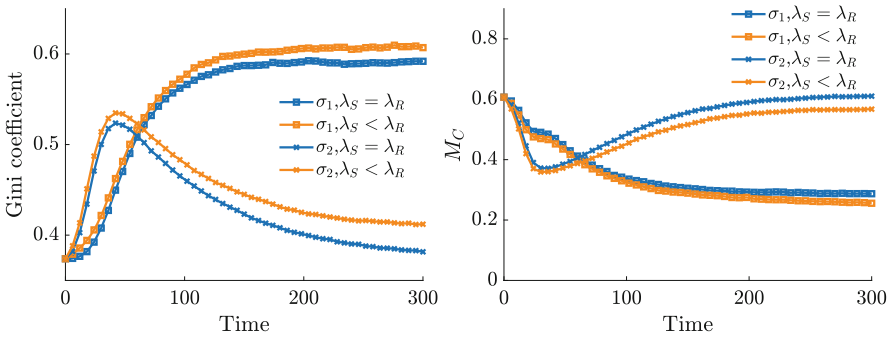


Fig. 3 **Test 2**. Behavior of the Gini index (left) and of the middle class fraction (right) defined in (46) during the outbreak of the epidemic for the different risk measures in (43) with $\alpha = 5$, $\sigma_0 = 0.1$

where $\alpha > 0$, $\sigma_0 > 0$. In detail, $\sigma_1(t)$ characterizes the instantaneous influence of the epidemic based on the observed number of infected, whereas $\sigma_2(t)$ takes into account possible long time memory effects on the market based on the epidemic impact.

We consider, as initial distribution, an inverse Gamma distribution

$$f(w) = \frac{(\mu-1)^\mu}{\Gamma(\mu)} \frac{\exp\left(-\frac{\mu-1}{w}\right)}{w^{1+\mu}} \tag{44}$$

with $\mu = 3$, representing an initial economic equilibrium state.

To get a more detailed view of the emerging equilibria, we resort to the Gini index calculation; see [48]. This value should be understood as a measure of a country's wealth inequality and varies in [0, 1], where 0 indicates perfect equality and 1 maximum inequality.

In Fig. 3 we represent the evolution of the Gini index. We clearly observe an inequality of wealth that grows with the epidemiological dynamics. Moreover, even in the case of σ_1 with $\lambda_S = \lambda_R$, where these effects are absorbed in the long-lasting trends, the recovery of the economy occurs at a much lower rate than the worsening rate.

Epidemiological dynamics may translate into additional wealth inequalities; in particular, we can measure the evolution of the total number of individuals belonging to the middle class. Although there are several ways to give a technical definition of the middle class, it is often more of an idea or estimate than a fixed number. Generally speaking, the middle class is loosely defined as those who fall into the middle group of workers compared to the bottom 20% or top 20%. We can define it using an interval $[w_L, w_R]$ such that

$$\int_0^{w_L} f(w,0) \approx 0.2, \qquad \int_{w_R}^{\infty} f(w,0) \approx 0.2, \tag{45}$$

and computing the time evolution of

$$M_C(t) = \int_{w_L}^{w_R} f(w,t)\,dw \tag{46}$$

gives us an estimate of the percentage of people living in middle-income households. In Fig. 3 (right plot), we represent the evolution of $M_C(t)$ corresponding to the considered $\sigma_1(t)$, $\sigma_2(t)$. We can clearly see how the emerging inequalities mainly affect the middle class, which is constantly decreasing in the case of σ_2 and undergoes a transitory decrease for σ_1. In particular, in this last scenario and in the $\lambda_S < \lambda_R$ regime, at the end of the epidemic dynamics only a partial recovery to the original pre-epidemic level is observed.

3 Social Control and Data Uncertainty

The adoption of containment measures to reduce the amplitude of the epidemic peak is a key aspect in tackling the rapid spread of an epidemic. Classical compartmental models need to be modified and studied to correctly describe the effects of forced external actions to reduce the impact of the disease. The importance of the social structure, such as age dependence, which was shown to be essential in the recent COVID-19 pandemic, must be considered. In addition, available data are often incomplete and heterogeneous, so a high degree of uncertainty must be incorporated into the model. In this section we deal with both these aspects following [8, 9].

3.1 Control of Socially Structured Models

The heterogeneity of the social structure that impacts the diffusion of the infective disease has been already discussed in Sect. 2. Among the social characteristics of fundamental importance in the context of the COVID-19 pandemic certainly the age of individuals is among the most significant given the heterogeneity of the contagiousness of the virus and the related health risks. We assume that the rapid spread of the disease and the low mortality rate allow to ignore changes in the social structure, such as the aging process, births, and deaths. This is equivalent to assuming in (4) that the interaction operators $Q_J(f_S; f_I; f_R)$, $J \in \{S, I, R\}$, vanish.

In order to keep to a standard notation in age-structured models [65, 66] in the following we will use the variable a to denote the social feature instead of w used in the general formulation (4). Thus, we will denote by $f_S(a,t)$, $f_I(a,t)$, and $f_R(a,t)$,

the distributions at time $t > 0$ of susceptible, infected, and recovered individuals with a given age $a \in \Lambda \subset \mathbb{R}^+$. In this situation the nonlinear incidence rate (5) reads

$$K(f_S, f_I)(a,t) = f_S(a,t) \int_\Lambda \beta_{\text{social}}(a, a_*) f_I(a_*, t)\, da_*, \tag{47}$$

where the function $\beta_{\text{social}}(a, a_*) \geq 0$ represents the interaction rate among individuals with different ages.

Typically, in socially structured models the interaction rate between people is assumed to be separable, and proportionate to the activity level of the social feature [64, 65], alternative approaches are based on preferential mixing [31, 60]. In what follows we will assume an age-dependent social interaction function such that

$$\beta_{\text{social}}(a, a_*) = \sum_{j \in \mathcal{A}} \beta_j(a, a), \tag{48}$$

where the set $\mathcal{A}$ indicates the social activities, such as family, work, school.

In the following, although we will derive our feedback controlled formulation for an age-structured SIR model, the extension to more realistic compartmental models [49, 57] and other social characteristics, such as the total number of social contacts and the wealth of individuals discussed in Sect. 2, can be done in a similar fashion.

3.1.1 Optimal Control Formulation

We consider an optimal control framework to define the strategy of a policy maker in introducing non-pharmaceutical interventions such as social distancing and other containment measures linked to the social structure. In what follows we assume that policy maker aims to minimize the impact of the epidemic through an age-dependent control action depending both on time and pairwise interactions among individuals with different ages. Thus, we introduce the optimal control problem

$$\begin{aligned} \min_{\mathbf{u} \in \mathcal{U}} J(\mathbf{u}) := & \int_0^T \psi(S(t), I(t)) dt \\ & + \sum_{j \in \mathcal{A}} \frac{1}{2} \int_0^T \int_{\Lambda \times \Lambda} \nu_j(a, a_*, t) |u_j(a, a_*, t)|^2 \, da da_* dt, \end{aligned} \tag{49}$$

subject to

$$\begin{aligned}
\frac{\partial f_S(a,t)}{\partial t} &= -f_S(a,t)\sum_{j\in\mathcal{A}}\int_\Lambda (\beta_j(a,a_*)-u_j(a,a_*,t))f_I(a_*,t)\,da_*,\\
\frac{\partial f_I(a,t)}{\partial t} &= f_S(a,t)\sum_{j\in\mathcal{A}}\int_\Lambda (\beta_j(a,a_*)-u_j(a,a_*,t))f_I(a_*,t)\,da_*\\
&\qquad -\gamma(a)f_I(a,t),\\
\frac{\partial f_R(a,t)}{\partial t} &= \gamma(a)f_I(a,t),
\end{aligned} \tag{50}$$

with initial condition $f_S(a,0)=f_S^0(a)$, $f_I(a,0)=f_I^0(a)$, and $f_R(a,0)=f_R^0(a)$. The function $\psi(S,I)$ accounts for the total number of the infected population $I(t)$ and susceptibles $S(t)$, such that $\psi(\cdot,\cdot)$ is positive and $\partial_I\psi(S,I)\geq 0$. This function models the policy maker's perception of the impact of the epidemic by the number of people currently infected and susceptible and in the sequel will be referred to as *perception function*. Each component of the control $\mathbf{u}=(u_1,\dots,u_L)$ acts selectively on the interaction between individuals of ages a and a_* for a specific activity in $\mathcal{A}$, with L the total number of activities. We consider a quadratic penalization of the control, weighted by a specific function $v_j(a,t)>0$ associated to each activity.

In (49) the set $\mathcal{U}\subseteq\mathbb{R}^L$ is the space of admissible controls u_j, $j\in\mathcal{A}$, defined as

$$\mathcal{U}=\left\{\mathbf{u}\in\mathbb{R}^L \,|\, 0\leq \mathcal{I}(u_j)(a,t)\leq \min\{M,\mathcal{I}(\beta_j)(a,t)\},\ \forall\,(a,t),\ M>0\right\},$$

where $\mathcal{I}$ corresponds to the integral operator

$$\mathcal{I}(\varphi)(a,t)=\frac{1}{I(t)}\int_\Lambda \varphi(a,a_*,t)f_I(a_*,t)\,da_*, \tag{51}$$

which ensures the admissibility of the solution for (50). The above restriction on admissible controls can be relaxed if we consider controls that violate the previous condition locally but preserve the inequality in integral form after integration against $i(a_*,t)$.

Solving the above optimization problem, however, is generally quite complicated and computationally demanding when there are uncertainties as it involves solving simultaneously the forward problem (49)–(50) and the backward problem derived from the optimality conditions [8]. Furthermore, the assumption that the policy maker follows an optimal strategy over a long time horizon seems rather unrealistic in the case of a rapidly spreading disease such as the COVID-19 epidemic.

Examples of Perception Function
We report two relevant examples of the perception function $\psi(\cdot)$, given by a convex function underestimating the number of infected

$$\psi(S, I)(t) = C\frac{I^q(t)}{q}, \qquad q \geq 1, \tag{52}$$

and a concave function overestimating such number

$$\psi(S, I)(t) = C\frac{\ln(1+\tau I(t))}{\tau S(t)}, \quad \tau > 0, \tag{53}$$

with $C > 0$ a suitable renormalization constant. The function in (52) has been introduced [8], whereas the function in (53) is related to well-known epidemic models with saturated incidence rates [30, 53, 76].

Let us emphasize that extending the above optimal control formulation to more complex compartmental models designed specifically for COVID-19 (see [57, 59]) can be done by generalizing the perception function in (49) to include, for example, the hospitalized compartment, or other specific indicators that can be measured from the data.

3.1.2 Feedback Controlled Compartmental Models

In this section we consider short time horizon strategies that permit to derive suitable feedback controlled models. These strategies are suboptimal with respect to the original problem (49)–(50), but they have proved to be very successful in several social modelling problems [2–5]. To this aim, we consider a short time horizon of length $h > 0$ and formulate a time discretized optimal control problem through the functional $J_h(u)$ restricted to the interval $[t, t+h]$, as follows:

$$\min_{\mathbf{u}\in\mathcal{U}} J_h(\mathbf{u}) := \psi(S(t), I(t+h)) + \sum_{j\in\mathcal{A}} \frac{1}{2}\int_{\Lambda\times\Lambda} v_j(a, a_*, t)|u_j(a, a_*, t)|^2 da da_* \tag{54}$$

subject to dynamics (50). By recalling that the macroscopic information on the infected is

$$I(t+h) = I(t) + h\int_\Lambda \Big[f_S(a,t)\sum_{j\in\mathcal{A}}\int_\Lambda \big(\beta_j(a,a_*) - u_j(a,a_*,t)\big) f_I(a_*,t) da_* - \gamma(a) f_I(a,t)\Big] da,$$

we can derive the minimizer of J_h computing $\nabla_{\mathbf{u}} J_h(\mathbf{u}) \equiv 0$. Using (54) and the macroscopic information on $I(t+h)$ and introducing the scaling $\nu_j(a, a_*, t) = h\kappa_j(a, a_*, t)$, we retrieve the instantaneous control

$$u_j(a, a_*, t) = \frac{1}{\kappa_j(a, a_*)} f_S(a, t) i(\cdot, a_*, t) \partial_I \psi(S(t), I(t+h))]. \tag{55}$$

Passing to the limit for $h \to 0$ and embedding into (50) the control u_j, we obtain an instantaneous feedback controlled dynamics.

Explicit Form of Incidence Rates

To understand the action of the feedback control (55), let us consider the simplest case of a standard SIR model without age dependence (homogeneous mixing), and specific social interactions. In this simplified setting the model has the structure of SIR model with the modified transmission rate

$$\begin{aligned} \frac{d}{dt} S(t) &= -\beta_\kappa(t) S(t) I(t), \\ \frac{d}{dt} I(t) &= \beta_\kappa(t) S(t) I(t) - \gamma I(t), \end{aligned} \tag{56}$$

where the transmission rate is

$$\beta_\kappa(t) = \beta - \frac{S(t) I(t) \partial_I \psi(S(t), I(t))}{\kappa}. \tag{57}$$

Introducing the explicit expressions of the control term for the perception function (52) and (53), we obtain in the convex case (52)

$$\beta_\kappa(t) = \beta - \frac{C S(t) I(t)^q}{\kappa} = \beta \left(1 - \frac{S(t) I(t)^q}{\kappa}\right), \tag{58}$$

whereas in the logarithmic case (53) and assuming $C = \beta$ and $\tau = 1/\kappa$, we have

$$\beta_\kappa(t) = \beta - \frac{C I(t)}{\kappa(1 + \tau I(t))} = \frac{\beta}{1 + \tau I(t)}. \tag{59}$$

Interestingly enough, the resulting nonlinear incidence rates (58)–(59) embedding the action of feedback controls correspond to the ones considered in [9, 80] and [30, 53], respectively. See also Sect. 2.1 of the present survey for a derivation of saturated incidence rates like (59). Other nonlinear incidence rates may be obtained similarly by considering different perception functions;

(continued)

see [80] and the references therein. We refer also to Sect. 2.1 for a different derivation of a saturated incidence function of the type (59).

Extensions to SEIRD Models

We can extend the previous computation to the socially structured compartmental model including additional compartments such as exposed, and dead individuals. The resulting feedback controlled SEIRD model reads

$$\begin{aligned}
\frac{\partial f_S(a,t)}{\partial t} &= f_S(a,t)\sum_{j\in\mathcal{A}}\int_\Lambda \big(\beta_j(a,a_*) - u_j(a,a_*,t)\big) f_I(a_*,t)da_*,\\
\frac{\partial f_E(a,t)}{\partial t} &= f_S(a,t)\sum_{j\in\mathcal{A}}\int_\Lambda \big(\beta_j(a,a_*) - u_j(a,a_*,t)\big) f_I(a_*,t)da_*,\\
&\qquad - \sigma(a) f_E(a,t),\\
\frac{\partial f_I(a,t)}{\partial t} &= \sigma(a) f_E(a,t+h) - (\gamma(a)+\alpha(a)) f_I(a,t),\\
\frac{\partial f_R(a,t)}{\partial t} &= \gamma(a) f_I(a,t),\\
\frac{\partial f_D(a,t)}{\partial t} &= \alpha(a) f_I(a,t),
\end{aligned} \tag{60}$$

with initial condition $f_S(a,0) = f_S^0(a)$, $f_E(a,0) = f_E^0(a)$, $f_I(a,0) = f_I^0(a)$, $f_R(a,0) = f_R^0(a)$, and $f_D(a,0) = f_D^0(a)$. Compared to (50) we introduced the age-dependent parameters: $\sigma(a) \geq 0$, the transition rate of exposed individuals to the infected class, and $\alpha(a) \geq 0$, the disease-induced death rate of infectious individuals. The feedback control u_j in this case is defined as follows:

$$u_j(a,a_*,t) = \frac{\sigma(a)}{\kappa_j(a,a_*)} f_S(a,t) f_I(a_*,t)\partial_I\psi(S(t),I(t)), \tag{61}$$

where the main difference with respect to (55) is the additional scaling parameter $\sigma(a)$. We refer to [8] for the derivation of the control form (61), and to [9] for further extension to SEPIAR model and an extensive study on $\psi(S,I) = I^q/q$.

3.1.3 Containment in Homogeneous Social Mixing Dynamics

To illustrate the effects of introduced controls that mimic containment procedures, let us first consider the case where the social structure is not present. Hence we consider model (56) with initial small number of infected and recovered $I(0) = 3.68 \times 10^{-6}$, $R(0) = 8.33 \times 10^{-8}$. These normalized fractions refer specifically to the first reported values in the case of the Italian outbreak of COVID-19, even if in this simple test case we will not try to match the data in a quantitative setting but simply to illustrate the behavior of the feedback controlled model.

Based on recent studies [81, 126], the initial infection rate of COVID-19 $R_0 = \beta/\gamma$ has been estimated between 2 and 6.5. Here, to exemplify the possible evolution of the pandemic we consider a value close to the lower bound, taking $\beta = 0.25$ and $\gamma = 0.10$, namely a recovery rate of 10 days, so that $R_0 = 2.5$.

In Fig. 4 we report the infected and recovered dynamics based on the activation of the control in two different time frames. Top images show the case in which the activation time is bounded $t \in [50, 100]$, which means that after 100 days we suppose that all containment restrictions are cancelled. In bottom row we consider a

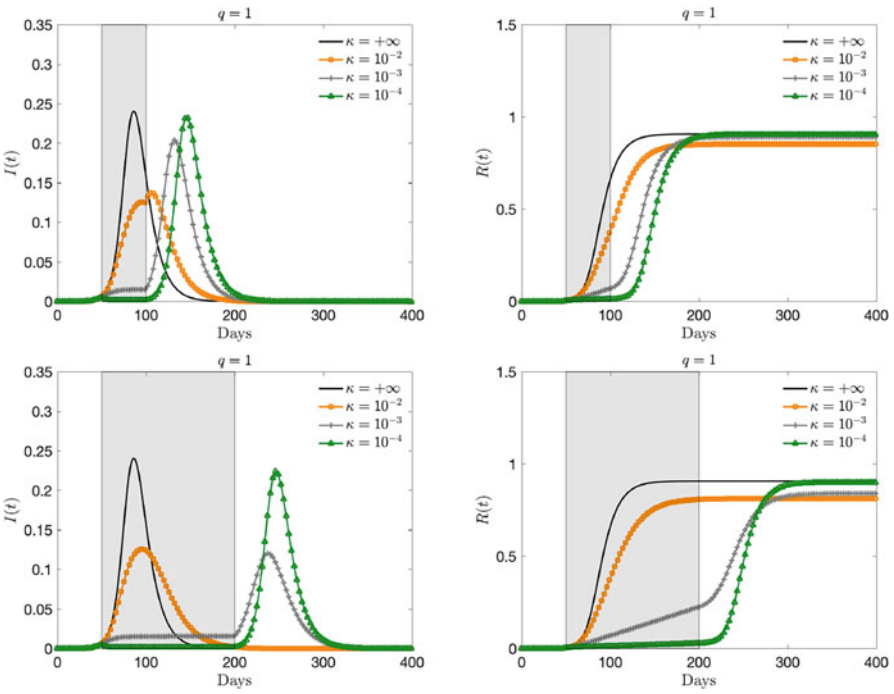


Fig. 4 Evolution of the fraction of infected (left) and recovered (right) based on the SIR feedback constrained model (56) with perception function $\psi(I) = I$ and several penalizations $\kappa = 10^{-2}, 10^{-3}, 10^{-4}$. The choice $\kappa = +\infty$ corresponds to the unconstrained case. First row, the control is applied for $t \in [50, 100]$. Second row, the control has a longer action in [50, 200]

larger activation time frame $t \in [50, 200]$. With the choice of the perception function $\psi(I) = I$ we can observe how the control term is able to flatten the curve.

3.2 Dealing with Data Uncertainty

Early in the outbreak of new infectious diseases, the actual number of people infected and recovered is typically underestimated, causing fatal delays in implementing public health policies in the face of spreading epidemic fronts. This is the case of the spreading of COVID-19 worldwide, often mistakenly underestimated due to deficiencies in surveillance and diagnostic capacity [89, 104, 126]. Among the common sources of uncertainties for dynamical systems modelling epidemic outbreaks we may consider: noisy and incomplete available data, and structural uncertainty due to the possible inadequacy of the mathematical model used to describe the phenomena under consideration. In the following we consider the effects on the dynamics of uncertain data, such as the initial conditions on the number of infected people or the interaction and recovery rates. On the numerical level we consider techniques based on stochastic Galerkin methods, for which spectral convergence on random variables is obtained under appropriate regularity assumptions [122].

3.2.1 Feedback Controlled and Socially Structured Models with Uncertain Inputs

We introduce the random vector $\mathbf{z} = (z_1, \dots, z_{d_z})$ whose components are assumed to be independent real valued random variables $z_k : (\Omega, F) \to (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, $k = 1, \dots, d_z$, with $\mathcal{B}_{\mathbb{R}}$ the Borel set. We assume to know the probability density $p(\mathbf{z}) : \mathbb{R}^{d_z} \to \mathbb{R}_+^{d_z}$ characterizing the distribution of $\mathbf{z}$. Here, $\mathbf{z} \in \mathbb{R}^{d_z}$ is a random vector taking into account various possible sources of uncertainty in the model.

In the presence of uncertainties we generalize the initial modelling by introducing the quantities $f_S(\mathbf{z}, a, t)$, $f_I(\mathbf{z}, a, t)$, and $f_R(\mathbf{z}, a, t)$ representing the distributions at time $t \geq 0$ of susceptible, infectious, and recovered individuals, respectively. The total size of the population is a deterministic conserved quantity in time, i.e.,

$$f_S(\mathbf{z}, a, t) + f_I(\mathbf{z}, a, t) + f_R(\mathbf{z}, a, t) = p(a), \qquad \int_\Lambda p(a)da = 1,$$

and the uncertain fractions of the population that are susceptible, infected, and recovered are defined as follows:

$$S(\mathbf{z}, t) = \int_\Lambda f_S(\mathbf{z}, a, t)\, da, \quad I(\mathbf{z}, t) = \int_\Lambda f_I(\mathbf{z}, a, t)\, da,$$

$$R(\mathbf{z},t)=\int_\Lambda f_R(\mathbf{z},a,t)\,da.$$

Hence, the controlled system (50) in the presence of uncertainty reads

$$\begin{aligned}
\partial_t f_S(\mathbf{z},a,t) &= -f_S(\mathbf{z},a,t)\sum_{j\in\mathcal{A}}\int_\Lambda \big(\beta_j(\mathbf{z},a,a_*)-u_j(a,a_*,t)\big)\,f_I(\mathbf{z},a_*,t)da_*,\\
\partial_t f_I(\mathbf{z},a,t) &= f_S(\mathbf{z},a,t)\sum_{j\in\mathcal{A}}\int_\Lambda \big(\beta_j(\mathbf{z},a,a_*)-u_j(a,a_*,t)\big)\,f_I(\mathbf{z},a_*,t)da_*\\
&\quad -\gamma(\mathbf{z},a))\,f_I(\mathbf{z},a,t),\\
\partial_t f_R(\mathbf{z},a,t) &= \gamma(\mathbf{z},a)\,f_I(\mathbf{z},a,t),
\end{aligned}\tag{62}$$

where the controls terms are assumed to be deterministic and defined as

$$u_j(a,a_*,t)=\frac{1}{\kappa_j(a,a_*)}\mathcal{R}[f_S(\cdot,a,t)f_I(\cdot,a_*,t)\partial_I\psi(S(\cdot,t),I(\cdot,t))],$$

being $\mathcal{R}[\psi(S(\cdot,t),I(\cdot,t))]$ a suitable operator taking into account the presence of the uncertainties $\mathbf{z}$. Examples of such operator that are of interest in epidemic modelling rely on the expected value

$$\mathcal{R}[\psi(S,I)(\cdot,t)]=\mathbb{E}[\psi(S,I)(\cdot,t)]=\int_{\mathbb{R}^{d_z}}\psi(S,I)(\mathbf{z},t)\;p(\mathbf{z})d\mathbf{z},\tag{63}$$

or on deterministic data that underestimate the number of infected

$$\mathcal{R}[\psi(S,I)(\cdot,t)]=\psi(S,I)(\mathbf{z}_0,t),\tag{64}$$

where $\mathbf{z}_0$ is a given value such that $I(\mathbf{z}_0,t)\le I(\mathbf{z},t)$, for all $\mathbf{z}\in\mathbb{R}^{d_z}$ and $t>0$. We refer to [8, 9] for further details on the derivation, and further extensions.

A Solvable Example

We consider a simplified version of model (62) in the absence of control, with homogeneous mixing $\beta_{\text{social}}(z)$ and recovery rate $\gamma(z)$, $z\in\mathbb{R}$, distributed as $p(z)$. Integrating against a we obtain the following SIR model with uncertainty:

$$\begin{aligned}
\frac{d}{dt}S(z,t) &= -\beta_{\text{social}}(z)S(z,t)I(z,t)\\
\frac{d}{dt}I(z,t) &= \beta_{\text{social}}(z)S(z,t)I(z,t)-\gamma(z)I(z,t),
\end{aligned}\tag{65}$$

(continued)

with deterministic initial values $I(z,0) = I_0$ and $S(z,0) = S_0$. Following [106], we assume a linear source of uncertainty $\beta_{\text{social}}(z) = \beta + \alpha z$, $\alpha > 0$, and constant recovery rate $\gamma(z) = \gamma > 0$. The solution for the proportion of infected during the initial exponential phase is

$$I(z,t) = I_0 e^{(\beta+\alpha z)S_0 t - \gamma t},$$

and its expectation

$$\mathbb{E}[I(\cdot,t)] = I_0 e^{\beta S_0 t - \gamma t} \int_{\mathbb{R}} e^{\alpha z S_0 t} p(z)\, dz = I_0 e^{\beta S_0 t - \gamma t} W(t), \tag{66}$$

where $W(t)$ represents the statistical correction factor to the standard deterministic exponential phase of the disease $I_0 e^{\beta S_0 t - \gamma t}$. If z is uniformly distributed in $[-1, 1]$, we can explicitly compute

$$W(t) = \frac{\sinh(\alpha S_0 t)}{\alpha S_0 t} > 1, \quad t > 0.$$

More in general, if z has zero mean, then by Jensen's inequality we have $W(t) > 1$ for $t > 0$, so that the expected exponential phase is amplified by the uncertainty.

In a similar way, keeping $\beta_{\text{social}}(z) = \beta$ constant, but introducing a source of uncertainty in the initial data $I(z,0) = I_0 + \mu z$, $\mu > 0$ and $z \in \mathbb{R}$ distributed as $p(z)$ the solution in the exponential phase reads

$$I(z,t) = (I_0 + \mu z) e^{\beta S_0 t - \gamma t},$$

and then its expectation

$$\mathbb{E}[I(\cdot,t)] = \int_{\mathbb{R}} (I_0 + \mu z) e^{\beta S_0 t - \gamma t} p(z)\, dz = (I_0 + \mu \bar{z}) e^{\beta S_0 t - \gamma t}, \tag{67}$$

where $\bar{z}$ is the mean of the variable z. Therefore, the expected initial exponential growth behaves as the one with deterministic initial data $I_0 + \mu \bar{z}$.

The presence of a large number of undetected infected is at the basis of the construction of numerous epidemiological models with an increasingly complex compartmental structure in which the original compartment of the infected is subdivided into further compartments with different roles in the propagation of the disease [51, 57, 59]. The following remark clarifies the relationships to other deterministic compartmental models.

Connection to Other Compartmental Models
Let us consider model (65) with a one-dimensional random input $z \in \mathbb{R}$ distributed as $p(z)$. Furthermore, for a function $F(z,t)$ we will denote its expected value as $\bar{F}(t) = \mathbb{E}[F(\cdot,t)]$. Now, starting from a discrete probability density function

$$p_k = P\{Z = z_k\}, \qquad \sum_{k=1}^{n} p_k = 1,$$

we have $\bar{F}(t) = \sum_{k=1}^{n} p_k F_k$, with $F_k = F(z_k)$. Taking the expectation in (65), we can write

$$\begin{aligned}
\frac{d}{dt}\bar{S}(t) &= -\bar{S}(t)\sum_{k=1}^{n}\tilde{\beta}_k p_k I_k(t),\\
\frac{d}{dt}\bar{I}(t) &= \bar{S}(t)\sum_{k=1}^{n}\tilde{\beta}_k p_k I_k(t) - \sum_{k=1}^{n}\gamma_k p_k I_k(t),\\
\frac{d}{dt}\bar{R}(t) &= \sum_{k=1}^{n}\gamma_k p_k I_k(t),
\end{aligned} \tag{68}$$

with $\tilde{\beta}_k = S_k\beta_k/\bar{S}$, $k = 1,\dots,n$. For example, in the case $n = 2$, by identifying $I_d = p_1 I_1$ and $I_u = p_2 I_2$ with the compartments of detected and undetected infectious individuals, we have the same structure of a SIAR compartmental model including the undetected (or the asymptomatic) class.

The additional dependence of the epidemiological parameters on the random variable allows us to take into account changes in the corresponding dynamics of disease transmission and recovery.

3.2.2 Application to the COVID-19 Outbreak

In this section, we first present the impact of social structure in feedback controlled models with uncertain data, which account for the presence of symptomatic and asymptomatic unreported cases, at the first wave of the COVID-19 pandemic.

Model Calibration and Estimating Actual Infection Trends
Estimating epidemiological parameters is a very difficult problem that can be addressed with different approaches [29, 33, 106]. In the case of COVID-19 due to the limited number of data and their great heterogeneity this becomes an even bigger problem that can easily lead to unrealistic results.

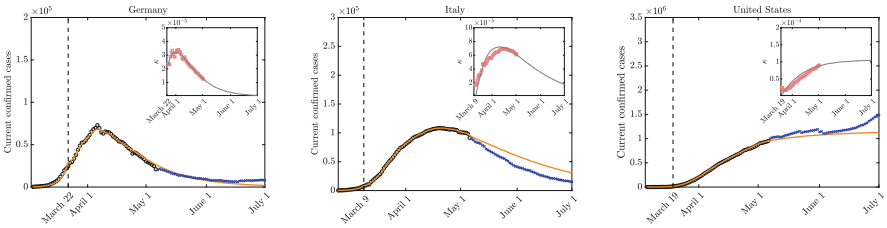


Fig. 5 Model behavior with fitting parameters and actual trends in the number of reported infectious using the estimated control penalization terms after lockdown over time in the various countries

Similarly to Sect. 2.1.5 we calibrate the model using data publicly available from the John Hopkins University GitHub repository, and GitHub repository of the Italian Civil Protection Department for the Italian case (see Data Sources in Sect. 5.1). We adopted the fitting procedure described in [9] that is based on a strategy with two optimization horizons (pre-lockdown and lockdown time spans) depending on the different strategies enacted by the governments of the considered European countries. Once the relevant epidemiological parameters have been estimated in the pre-lockdown time span, i.e., $\beta_e > 0$ and $\gamma_e > 0$, we successively proceeded with the estimation of the control penalty parameter $\kappa_e = \kappa(t) > 0$. These two calibration steps were analyzed under the assumption of homogeneous mixing.

We report in Fig. 5 the corresponding time dependent values for the control parameter $k(t)$, as well as results of the model fitting with the actual trends of infected individuals.

Next we focus on the influence of uncertain quantities on the controlled system with homogeneous mixing. According to recent results on the diffusion of COVID-19 in many countries the number of infected, and therefore recovered, is largely underestimated on the official reports, see, e.g., [67, 89]. As discussed in [9], a parameter estimation based on the previous fitting has some limitations and in particular overestimates the reproduction rate in the early phase of the pandemic. For this purpose, to have an insight on global impact of uncertain parameters we consider a two-dimensional uncertainty $\mathbf{z} = (z_1, z_2)$ with independent components such that

$$I(\mathbf{z}, 0) = I_0(1 + \mu z_1), \qquad R(\mathbf{z}, 0) = R_0(1 + \mu z_1), \qquad \mu > 0 \tag{69}$$

and

$$\beta(\mathbf{z}) = \beta_e - \alpha_\beta z_2, \qquad \gamma(\mathbf{z}) = \gamma_e + \alpha_\gamma z_2, \qquad \alpha_\beta, \alpha_\gamma > 0 \tag{70}$$

where z_1, z_2 are chosen to be distributed as symmetric Beta distributions in $[0, 1]$ and f_I^0 and f_R^0 are the initial number of reported cases and recovered taken from [126]. The parameter $\mu = 2(c - 1)$ is common for all countries such that $\mathbb{E}[I(\mathbf{z}, 0)] = cI(0)$, $\mathbb{E}[R(\mathbf{z}, 0)] = cR(0)$ where $c = 8.56$, corresponding to average

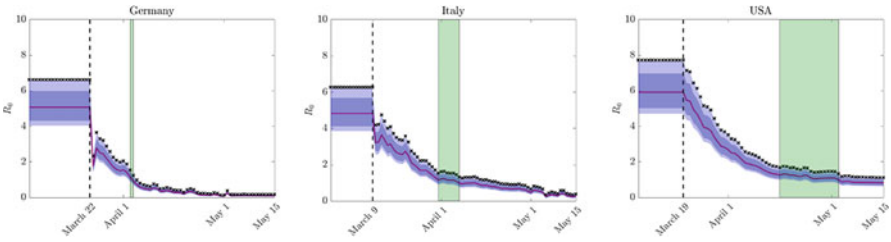


Fig. 6 Evolution of estimated reproduction number R_0 and its confidence bands for uncertain data as in (70). The 95 and 50% confidence levels are represented as shaded and darker shaded areas, respectively. The green zones denote the interval between the first day of the 50% confidence band and the expected value fall below 1

disagreement in the total number of cases based on an estimated infection fatality rate (IFR) of 1.3% in the range 0.9%–2.0%. The feedback controlled model has been computed using an estimation of the total number of susceptible and infected reported, namely we have the control term

$$u(t) = -\frac{1}{k(t)} S_r(t) I_r(t), \tag{71}$$

where $S_r(t)$ and $I_r(t)$ are the model solution obtained from the registered data, and thus $I_r(t)$ represents a lower bound for the uncertain solution $I(\mathbf{z}, t)$.

In Fig. 6 we report the evolution of reproduction number R_0 for the considered countries under the uncertainties in (70) obtained with $\alpha_\beta = 0.03$, $\alpha_\gamma = 0.05$, and $z_2 \sim B(2, 2)$. The reproduction number is estimated from

$$R_0(z_2, t) = \frac{\beta(z_2) - u(t)\chi(t > \bar{t})}{\gamma(z_2)},$$

being the control $u(t)$ defined in (71) and $\bar{t}$ is the country-dependent lockdown time. The estimated reproduction number relative to data is reported with x-marked symbols and represents an upper bound for $R_0(z_2, t)$.

Effect of Social Contacts in the Population

We first analyze the effects of the inclusion of age dependence and social interactions in the above dynamics with uncertainty in the case of COVID-19 outbreak for the Italian case. The age-dependent social interaction rate $\beta(a, a_*)$ is defined as follows:

$$\beta(a, a_*) = (1 - \xi)\beta_e + \xi \sum_{j \in \mathcal{A}} \beta_j(a, a_*), \tag{72}$$

where $0 \leq \xi \leq 1$; thus, for $\xi = 0$ we recover the homogeneous mixing, whereas for $\xi = 1$ we have a full social mixing behavior.

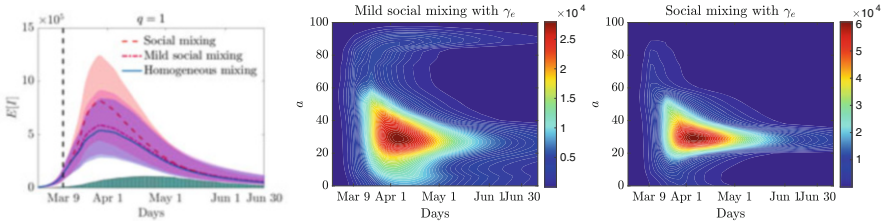


Fig. 7 Left: Expected number of infected in time for the perception function $\psi(I) = I$, and a constant recovery rate γ_e together with the confidence bands for homogeneous mixing ($\xi = 0$), mild social mixing ($\xi = 0.75$), and full social mixing ($\xi = 1$). Middle and right: Expected age distribution of infectious individuals with constant γ_e

The social interaction function, $\beta_{\text{social}}(a, a_*)$, accounts for the interactions due to specific activities $\mathcal{A} = \{\text{Family, Education, Profession}\}$. This function is normalized using the estimated parameters β_e in accordance with

$$\beta_e = C_\beta \int_{\Lambda\times\Lambda} \beta(a, a_*) f(a) f(a_*)\, da\, da_*, \quad \gamma_e = C_\gamma \int_\Lambda \gamma(a) f(a)\, da, \tag{73}$$

where $f(a)$ is the age distribution with $\Lambda = [0, a_{\max}]$, $a_{\max} = 100$, and C_β, C_γ normalization constants. We refer to the Appendix of [8] for specific definition of the social interaction $\beta_j(a, a_*)$.

In Fig. 7 we report the results of the expected number of infected with the related confidence bands in case of homogeneous mixing and different levels of social mixing ($\xi = 0.75$, $\xi = 1$) for the constant recovery rate γ_e. Middle and right figures report the corresponding expected density of infected individuals $f_I(a, t) = \mathbb{E}(f_I(z, a, t))$ for mild and full social mixing. Uncontrolled homogeneous mixing model is used in the pre-lockdown phase (before 9 March), whereas the feedback controlled age-dependent model (62) is used in the lockdown phase.

Relaxing Control on the Various Social Activities

We consider the social interaction functions corresponding to the contact matrices in [102] for the various countries. As a result, we have four interaction functions characterized by $\mathcal{A} = \{F, E, P, O\}$, where we identify family and home contacts with β_F, education and school contacts with β_E, professional and work contacts with β_P, and other contacts with β_O. We report in Fig. 8, as an example, the total social interaction functions for the various countries. The functions share a similar structure but with different scalings according to the country specific features.

An age-related recovery rate $\gamma(a)$ is selected according to [9] as a decreasing function of the age:

$$\gamma(a) = \gamma_e + Ce^{-ra}, \tag{74}$$

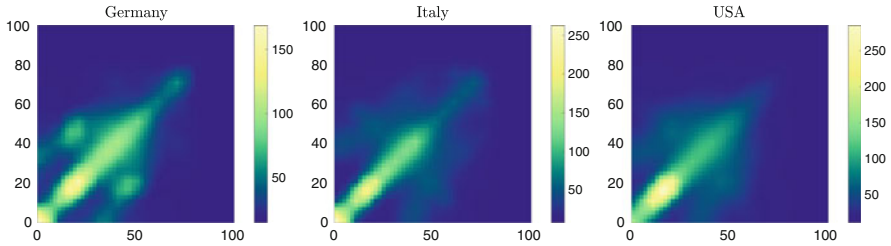


Fig. 8 The contact interaction function $\beta = \beta_F + \beta_E + \beta_P + \beta_O$ taking into account the contact rates of people with different ages. Family and home contacts are characterized by β_F, education and school contacts by β_E, professional and work contacts by β_P, and other contacts by β_O

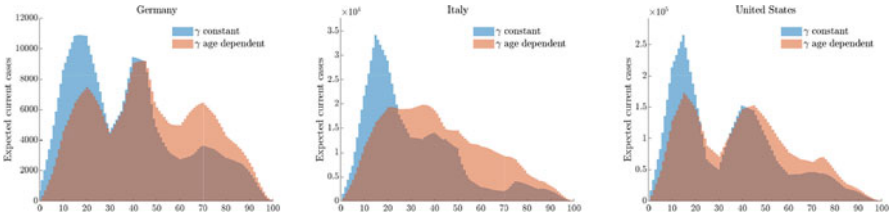


Fig. 9 Age distribution of infected using constant and age-dependent recovery rates as in (74) at the end of the lockdown period in different countries

with $r = 5$ and $C \in \mathbb{R}$, in accordance with [120, 124], and such that the normalization (73) is satisfied.

To match the single control applied in the extrapolation of the penalization term $\kappa(t)$ to age-dependent penalization factors $\kappa_j(a,t)$, we redistribute the values of the penalization parameters as

$$\kappa_j(a,t)^{-1} = \frac{w_j(t)\int_\Lambda \beta_j(a,a_*)\,da_*}{\sum_{j\in\mathcal{A}} w_j(t)\int_{\Lambda\times\Lambda}\beta_j(a,a_*)\,da\,da_*}\kappa(t)^{-1}, \quad j \in \mathcal{A},$$

where $w_j(t) \geq 0$ are weight factors denoting the relative amount of control on a specific activity. According to [102], we assume $w_E = 1.5$, $w_H = 0.2$, $w_P = 0.5$, and $w_O = 0.6$, namely the largest effort of the control is due to the school closure, which as a consequence implies more interactions at home. Work and other activities are equally impacted by the lockdown.

In Fig. 9 we report the age distribution of infected computed for each country at the end of the lockdown period using an age-dependent recovery and a constant recovery. The differences in the resulting age distributions are evident. In subsequent simulations, to avoid an unrealistic peak of infection among young people, we decided to adopt an age-dependent recovery [120].

We analyze the effects on each country of the same relaxation of the lockdown measures at two different times. The first date is country specific according to current available information, and the second is June 1st for all countries. For all

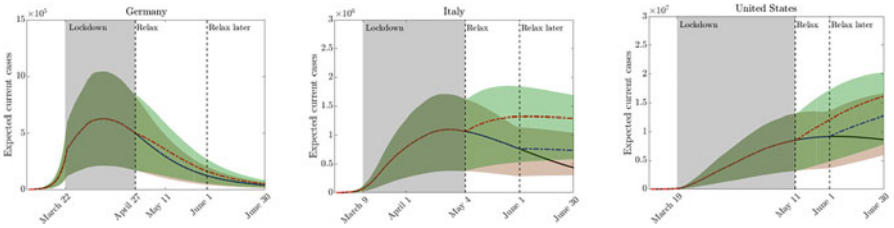


Fig. 10 Effect on releasing containment measures in various countries at two different times. In all countries after lockdown we assumed a reduction of individual controls on the different activities by 20% on family activities, 35% on work activities, and 30% on other activities by keeping the lockdown over the school

countries, we assumed a reduction of individual controls on the different activities by 20% on family activities, 35% on work activities, and 30% on other activities without changing the control over the school. The behaviors of the curves of infected people together with the relative 95% confidence bands are reported in Fig. 10.

The results show well the substantial differences between the different countries, with a situation in the USA, which highlight that the relaxation of lockdown measures could lead to a resurgence of the infection. On the contrary, Germany was in the most favorable situation to ease the lockdown without risking a new start of the infection.

4 Multiscale Transport Models

In this section, we introduce multiscale hyperbolic transport models designed to study the propagation of an epidemic phenomenon described by the diffusive behavior of the non-commuting part of the population, acting only over an urban scale, and the spatial movement and interaction of commuters, moving also on an extra-urban scale. This makes it possible to describe more realistically the typical dynamic of commuters, which affects only a small fraction of individuals, and to distinguish it from the epidemic process that, instead, involves the entire population, including non-commuters. The presence of a group of non-commuting population, indeed, prevents the entire population in a compartment from moving indiscriminately through space generating an unrealistic mass migration effect. In the following, we will consider a spatial domain either structured as a network, whose nodes identify cities of interest and arcs represent common mobility paths, as discussed in Sect. 4.1, or representing realistic 2D geographical regions, as further detailed in Sect. 4.2.

4.1 Spatial Dynamics on Networks

4.1.1 1D Hyperbolic Compartmental Model

To simplify the presentation, the epidemiological starting point of the model is given by a compartmental structure with a simple SIR partitioning [65, 75]. We assume to have a population with individuals having no prior immunity and the vital dynamics represented by births and deaths is neglected because of the time scale considered. To account for the spatial movement of the population, individuals of each compartment are subdivided into three classes, $S_{\pm,0}$, $I_{\pm,0}$, and $R_{\pm,0}$, traveling in a 1D bounded space domain $\Omega \subseteq \mathbb{R}$ with characteristic speeds $+\lambda_i$, $-\lambda_i$, and 0 respectively, with $i \in \{S, I, R\}$. Therefore, we consider a stationary part of the population, of non-commuters, characterized by a null characteristic speed. The total compartmental densities are defined as the sum of all the components of the subgroups

$$S = S_+ + S_- + S_0, \qquad I = I_+ + I_- + I_0, \qquad R = R_+ + R_- + R_0 \,. \tag{75}$$

The discrete-velocity system of the SIR epidemic transport model for commuters, associated to relaxation times τ_i, then reads

$$\begin{aligned}
\frac{\partial S_\pm}{\partial t} \pm \lambda_S \frac{\partial S_\pm}{\partial x} &= -F_I(S_\pm, I) + \frac{1}{2\tau_S}\left(S_\mp - S_\pm\right) \,, \\
\frac{\partial I_\pm}{\partial t} \pm \lambda_I \frac{\partial I_\pm}{\partial x} &= F_I(S_\pm, I) - \gamma_I I_\pm + \frac{1}{2\tau_I}\left(I_\mp - I_\pm\right) \,, \\
\frac{\partial R_\pm}{\partial t} \pm \lambda_R \frac{\partial R_\pm}{\partial x} &= \gamma_I I_\pm + \frac{1}{2\tau_R}\left(R_\mp - R_\pm\right) \,.
\end{aligned} \tag{76}$$

This system is coupled with a classical ODE SIR model, which describes the evolution of the stationary population of non-commuters:

$$\frac{\mathrm{d}S_0}{\mathrm{d}t} = -F_I(S_0, I) \,, \qquad \frac{\mathrm{d}I_0}{\mathrm{d}t} = F_I(S_0, I) - \gamma_I I_0 \,, \qquad \frac{\mathrm{d}R_0}{\mathrm{d}t} = \gamma_I I_0 \,. \tag{77}$$

Let us observe that, under no inflow/outflow boundary conditions, summing up the equations in (76)–(77) and integrating in Ω yields the conservation of the total population.

All the epidemic densities and, eventually, the epidemic parameters and the relaxation times depend on $(\mathbf{z}, x, t)$, where (x, t) are the physical variables of space $x \in \Omega \subseteq \mathbb{R}$ and time $t > 0$, while $\mathbf{z} = (z_1, \ldots, z_d)^T \in \mathbb{R}^d$ is a random vector characterizing the possible sources of uncertainty as in Sect. 3.2.1. The same applies for the incidence function F_I, defined with respect to the infectious compartment I as

$$F_I(g, I) = \beta_I \frac{g I^p}{1 + \kappa_I I^p}, \qquad p \geq 1, \tag{78}$$

where $\beta_I(\mathbf{z}, x, t)$ is the transmission rate, accounting for both number of contacts and probability of transmission, hence it may vary based on the effects of government control actions, such as mandatory wearing of masks, shutdown of specific work/school activities, or full lockdowns [8, 59, 65]. The parameter $\kappa_I(\mathbf{z}, x, t)$ acts as incidence damping coefficients based on the self-protective behavior of the individual that arises from awareness of the risk associated with the epidemic [18, 30, 53]. We refer also to Sects. 2.1 and 3 for the derivation of saturated incidence functions of the form (78). Note that the classic bilinear case corresponds to $p = 1$ and $\kappa_I = 0$. Finally, the parameter $\gamma_I(\mathbf{z}, x, t)$ is the recovery rate of infected (inverse of the infectious period).

4.1.2 Macroscopic Formulation and Diffusion Limit

Introducing now the macroscopic variables S_c, I_c, R_c for the commuters, with $S_c = S_+ + S_-$, $I_c = I_+ + I_-$, $R_c = R_+ + R_-$, and defining the fluxes

$$J_S = \lambda_S(S_+ - S_-), \quad J_I = \lambda_I(I_+ - I_-), \quad J_R = \lambda_R(R_+ - R_-), \tag{79}$$

a hyperbolic model underlying the macroscopic formulation of the spatial propagation of an epidemic at finite speeds, equivalent to the mesoscopic one [11], presented in system (76), is obtained [18]:

$$\begin{aligned}
\frac{\partial S_c}{\partial t} + \frac{\partial J_S}{\partial x} &= -F_I(S_c, I)\,, \\
\frac{\partial I_c}{\partial t} + \frac{\partial J_I}{\partial x} &= F_I(S_c, I) - \gamma_I I_c\,, \\
\frac{\partial R_c}{\partial t} + \frac{\partial J_R}{\partial x} &= \gamma_I I_c\,, \\
\frac{\partial J_S}{\partial t} + \lambda_S^2 \frac{\partial S_c}{\partial x} &= -F_I(J_S, I) - \frac{1}{\tau_S} J_S\,, \\
\frac{\partial J_I}{\partial t} + \lambda_I^2 \frac{\partial I_c}{\partial x} &= \frac{\lambda_I}{\lambda_S} F_I(J_S, I) - \gamma_I J_I - \frac{1}{\tau_I} J_I\,, \\
\frac{\partial J_R}{\partial t} + \lambda_R^2 \frac{\partial R_c}{\partial x} &= \frac{\lambda_R}{\lambda_I} \gamma_I J_I - \frac{1}{\tau_R} J_R\,.
\end{aligned} \tag{80}$$

Note that here the above system is coupled with the equations for the non-commuting population (77) through identities (75). It is easy to verify that system (80) is symmetric hyperbolic in the sense of Friedrichs–Lax [54].

From a formal viewpoint, it can be shown that the proposed model recovers the parabolic behavior expected from standard space-dependent epidemic models in the diffusion limit [12, 18]. Introducing the diffusion coefficients $D_i = \lambda_i^2 \tau_i, i \in \{S, I, R\}$, that characterize the diffusive transport mechanism of S, I, R, respectively, and letting $\tau_i \to 0$, while keeping the diffusion coefficients finite [79], from the last three equations of system (80) we recover Fick's laws

$$J_S = -D_S \frac{\partial}{\partial x} S_c, \quad J_I = -D_I \frac{\partial}{\partial x} I_c, \quad J_R = -D_R \frac{\partial}{\partial x} R_c,$$

which, inserted in the first three equations of system (80), yield the following parabolic reaction–diffusion system for the commuters [17, 90]:

$$\begin{aligned} \frac{\partial S_c}{\partial t} &= \frac{\partial}{\partial x}\left(D_S \frac{\partial}{\partial x} S_c \right) - F_I(S_c, I)\,, \\ \frac{\partial I_c}{\partial t} &= \frac{\partial}{\partial x}\left(D_I \frac{\partial}{\partial x} I_c \right) + F_I(S_c, I) - \gamma_I I_c\,, \\ \frac{\partial R_c}{\partial t} &= \frac{\partial}{\partial x}\left(D_R \frac{\partial}{\partial x} R_c \right) + \gamma_I I_c\,. \end{aligned} \tag{81}$$

The relaxation times can modify the nature of the behavior of the solution [12, 18], which can result either hyperbolic or parabolic (when considering small relaxation times and large speeds). This feature of the model makes it particularly suitable for the description of the dynamics of human populations, which are characterized by movement at different spatial scales [23]. It is therefore natural to assume $\tau_i = \tau_i(x)$, since in geographic areas that are densely populated we can assume a diffusive dynamics, while along the main arteries of communication a hyperbolic description will be more appropriate avoiding propagation of information at infinite speed.

Reproduction Number in Space Dependent Dynamics

The standard threshold of epidemic models is the well-known basic reproduction number R_0. Its definition in the case of spatially dependent dynamics, as already noted in [116, 117], is not straightforward particularly when considering its spatial dependence.

Assuming no inflow/outflow boundary conditions in Ω, summing up the evolutionary equations for the infectious compartment I in (76)–(77) and integrating over space we have

$$\frac{\partial}{\partial t} \int_\Omega I(\mathbf{z}, x, t)\, dx = \int_\Omega F_I(S, I)\, dx - \int_\Omega \gamma_I(\mathbf{z}, x, t) I(\mathbf{z}, x, t)\, dx \geq 0$$

(continued)

when

$$R_0(\mathbf{z}, t) = \frac{\int_\Omega F_I(S, I)\, dx}{\int_\Omega \gamma_I(\mathbf{z}, x, t) I(\mathbf{z}, x, t)\, dx} \geq 1, \tag{82}$$

which corresponds to definition (6) in the case of a social-dependent dynamics.

If no spatial dependence is assigned to variables and parameters, as well as no uncertainty, and no social distancing effects are taken into account, i.e., $\kappa_I = 0$, we recover the conventional SIR ODE model and the reproduction number results in accordance with its standard definition [65]:

$$R_0(t) = \frac{\beta_I S}{\gamma_I}.$$

4.1.3 Extension to Multi-Compartmental Modelling

To account for more complex compartmental models capable of better analyzing the evolution of specific infectious diseases, we consider extending the simple SIR compartmentalization by taking into account two additional population compartments, E and A, resulting in a SEIAR model [19, 20]. Subjects in the E compartment are the exposed, hence infected but not yet infectious, being in the latent period. Moreover, among the infectious subjects, we distinguish the population between a group of individuals I who will develop severe symptoms and a group of individuals A who will never develop symptoms or, if they do, these will be very mild. In fact, as discussed in Sect. 3, the presence of undetected asymptomatic individuals turns out to be essential to correctly analyze the evolution of COVID-19 [57, 98].

Note that the presence of uncertainty in the data, included from the beginning in the modelling process, could allow the compartmentalization of asymptomatic individuals to be eliminated by implicitly including them in the uncertainty about the number of infected individuals, as described in Sect. 3.2.1. In this context, however, in order to highlight the link with similar models used in the literature [98, 111, 112], we keep the asymptomatic compartment separated to the symptomatic one, with the former being affected by the highest level of uncertainty.

Defining the total density of the additional compartments, $E = E_+ + E_- + E_0$, $A = A_+ + A_- + A_0$, the resulting discrete-velocity system of the SEIAR epidemic transport model for commuters reads

$$
\begin{aligned}
\frac{\partial S_\pm}{\partial t} \pm \lambda_S \frac{\partial S_\pm}{\partial x} &= -F_I(S_\pm, I) - F_A(S_\pm, A) + \frac{1}{2\tau_S}(S_\mp - S_\pm)\,, \\
\frac{\partial E_\pm}{\partial t} \pm \lambda_E \frac{\partial E_\pm}{\partial x} &= F_I(S_\pm, I) + F_A(S_\pm, A) - aE_\pm + \frac{1}{2\tau_E}(E_\mp - E_\pm)\,, \\
\frac{\partial I_\pm}{\partial t} \pm \lambda_I \frac{\partial I_\pm}{\partial x} &= a\sigma E_\pm - \gamma_I I_\pm + \frac{1}{2\tau_I}(I_\mp - I_\pm)\,, \\
\frac{\partial A_\pm}{\partial t} \pm \lambda_A \frac{\partial A_\pm}{\partial x} &= a(1-\sigma)E_\pm - \gamma_A A_\pm + \frac{1}{2\tau_A}(A_\mp - A_\pm)\,, \\
\frac{\partial R_\pm}{\partial t} \pm \lambda_R \frac{\partial R_\pm}{\partial x} &= \gamma_I I_\pm + \gamma_A A_\pm + \frac{1}{2\tau_R}(R_\mp - R_\pm)\,,
\end{aligned}
\tag{83}
$$

which is coupled with the following SEIAR model describing the evolution of non-commuting individuals:

$$
\begin{aligned}
\frac{\mathrm{d}S_0}{\mathrm{d}t} &= -F_I(S_0, I) + F_A(S_0, A)\,, \\
\frac{\mathrm{d}E_0}{\mathrm{d}t} &= F_I(S_0, I) + F_A(S_0, A) - aE_0\,, \\
\frac{\mathrm{d}I_0}{\mathrm{d}t} &= a\sigma E_0 - \gamma_I I_0\,, \\
\frac{\mathrm{d}A_0}{\mathrm{d}t} &= a(1-\sigma)E_0 - \gamma_A A_0\,, \\
\frac{\mathrm{d}R_0}{\mathrm{d}t} &= \gamma_I I_0 + \gamma_A A_0\,.
\end{aligned}
\tag{84}
$$

The quantity $\gamma_A(\mathbf{z}, x, t)$ is the recovery rate of asymptomatic/mildly symptomatic infected, which is distinguished from the recovery rate of highly symptomatic infected previously introduced $\gamma_I(\mathbf{z}, x, t)$, while $a(\mathbf{z}, x, t)$ represents the inverse of the latency period and $\sigma(\mathbf{z}, x, t)$ is the probability rate of developing severe symptoms [28, 57, 112]. In this model, the transmission of the infection is governed by two different incidence functions, $F_I(\cdot, I)$ and $F_A(\cdot, A)$, simply to distinguish between the behavior of I and A individuals. Analogously to (78),

$$
F_A(g, A) = \beta_A \frac{gA^p}{1 + \kappa_A A}, \tag{85}
$$

where a different contact rate, β_A, and coefficient κ_A are taken into account for mildly/no symptomatic people. The flow chart of the multiscale SEIAR model is shown in Fig. 11.

Let us observe that, similarly to the diffusive scaling presented in Sect. 4.1.2 for the SIR-type model, introducing the same definition (79) of flux for the additional compartments, J_E and J_A, we get an analogous macroscopic formulation also for

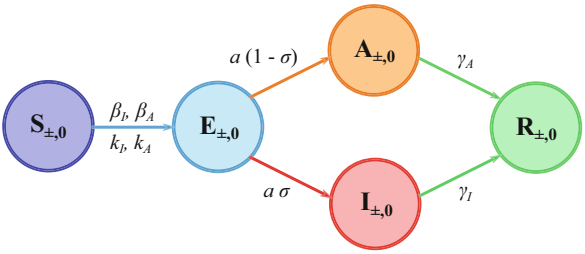


Fig. 11 Flow chart of the multi-population SEIAR dynamics based on five compartments: susceptible (S), exposed (E), severe symptomatic infectious (I), mildly symptomatic/asymptomatic infectious (A), and recovered—healed or deceased—population (R), each one subdivided into three classes of individuals traveling in the domain with characteristic speeds $+\lambda_i$, $-\lambda_i$, and 0, with $i \in \{S, E, I, A, R\}$

the SEIAR-type spatial model. Furthermore, defining also $D_E = \lambda_E^2 \tau_E$ and $D_A = \lambda_A^2 \tau_A$, we recover the diffusion limit of the SEIAR-type system. The reader can refer to [19] for details on this derivation.

Reproduction Number in Space Dependent SEIAR Models
For the SEIAR-type spatial model, the reproduction number (which is again not straightforward to be determined) can be computed following the *Next-Generation Matrix* (NGM) approach [41] considering no flux boundary conditions, which yields the following definition for the average value of R_0 in the domain Ω for $t > 0$, given the uncertain input vector $\mathbf{z}$:

$$R_0(\mathbf{z}, t) = \frac{\int_\Omega F_I(S, I)\, dx}{\int_\Omega \gamma_I(\mathbf{z}, x, t) I(\mathbf{z}, x, t)\, dx} \cdot \frac{\int_\Omega a(\mathbf{z}, x, t)\sigma(\mathbf{z}, x, t) E(\mathbf{z}, x, t)\, dx}{\int_\Omega a(\mathbf{z}, x, t) E(\mathbf{z}, x, t)\, dx} + \frac{\int_\Omega F_A(S, A)\, dx}{\int_\Omega \gamma_A(\mathbf{z}, x, t) A(\mathbf{z}, x, t)\, dx} \cdot \frac{\int_\Omega a(\mathbf{z}, x, t)(1 - \sigma(\mathbf{z}, x, t)) E(\mathbf{z}, x, t)\, dx}{\int_\Omega a(\mathbf{z}, x, t) E(\mathbf{z}, x, t)\, dx}. \tag{86}$$

We refer to [19] for the details of the derivation of the above expression.

4.1.4 Network Modelling

The hyperbolic transport models here proposed, similarly to other fields of application, like traffic flow models, chemotaxis, and cardiovascular modelling, can be embedded into a network of cities following [26, 100]. Note that the approach differs from the classical network modelling in epidemiology based on coupled systems of ODEs [13, 39, 57, 85].

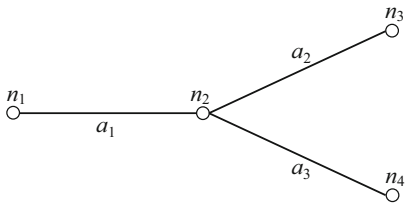


Fig. 12 Schematic representation of a network composed by 4 nodes (n_1, n_2, n_3, n_4) and 3 arcs (a_1, a_2, a_3) in a Y-shape configuration

A network or a connected graph $\mathcal{G} = (\mathcal{N}, \mathcal{A})$ is composed of a finite set of N nodes (or vertices) $\mathcal{N}$ and a finite set of A bidirectional arcs (or edges) $\mathcal{A}$, such that an arc connects a pair of nodes [100]. An example of network is presented in Fig. 12.

Following [18, 20], it is possible to structure a 1D network considering that the nodes of the network identify locations of interest such as municipalities, provinces or, in a wider scale, regions or nations, while the arcs, enclosing the 1D spatial dynamics, represent the paths linking each location to the others. In this configuration, nodes are *active* since the epidemic state of each one evolves in time influenced by the mobility of the commuting individuals, moving from the other locations included in the network, always considering a part of the population composed by non-commuting individuals that remain at the origin node.

In order to prescribe the proper coupling between nodes and arcs, it is necessary to impose appropriate transmission conditions at each arc–node interface, which ensure the conservation of total density (population) in the network and of fluxes at the interface and further solving the Riemann problem at each interface employing Riemann Invariants. The complete description of the implementation of transmission conditions at nodes is presented in [18] for a SIR-type spatial model and in [20] for a SEIAR-type transport model.

4.1.5 Effect of Spatially Heterogeneous Environments in Hyperbolic and Parabolic Configuration

Following [119], we analyze the behavior of the SIR-type model (76) with a commuter-only population ($i = i_c, i \in \{S, I, R\}$) in a single 1D domain concerning spatially heterogeneous environments, taking into account a spatially variable contact rate

$$\beta_I(x) = \hat{\beta}_I \left(1 + 0.05 \sin \frac{13\pi x}{20}\right).$$

Initial conditions are imposed assuming, in this setting, no uncertainty in the input data, with

$$S(x, 0) = 1 - I(x, 0), \qquad I(x, 0) = 0.01\, e^{-(x-10)^2}, \qquad R(x, 0) = 0.0,$$

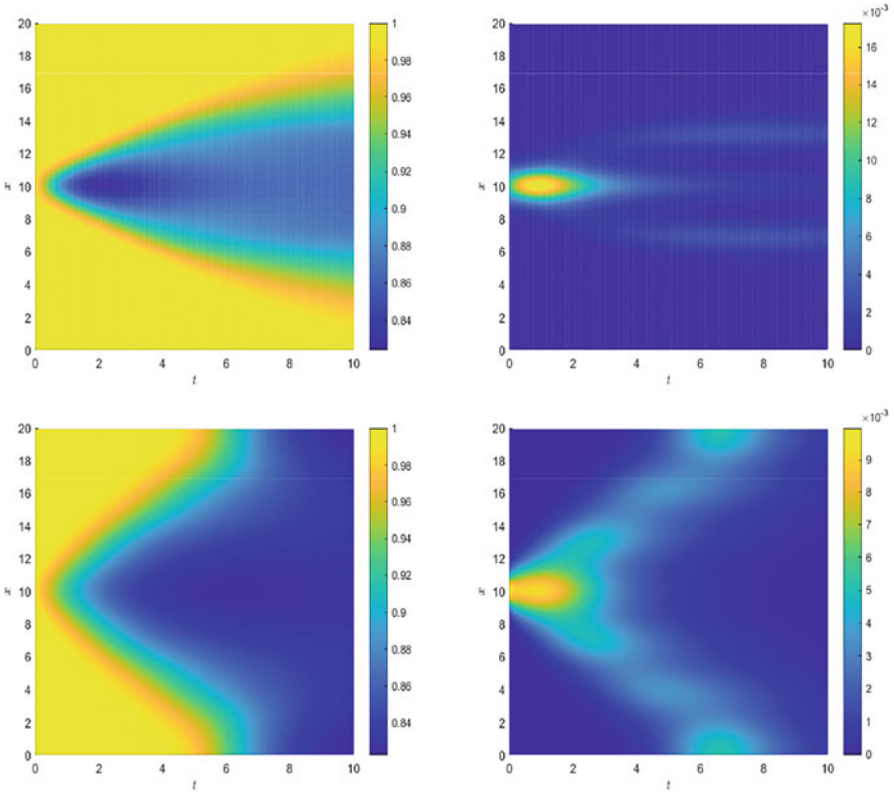


Fig. 13 Numerical results of the spatially heterogeneous SIR case with hyperbolic configuration of relaxation times and characteristic velocities (first row), with $\tau = 1.0, \lambda^2 = 1.0$, or parabolic configuration (second row), with $\tau = 10^{-5}, \lambda^2 = 10^5$. Time and spatial evolution of S (first column) and I (second column)

fluxes $J_S(x, 0) = J_I(x, 0) = J_R(x, 0) = 0.0$ and zero-flux boundary conditions. The initial reproduction number results $R_0 = 1.111 > 1$, given by the choice $\hat{\beta}_I = 11.0$, $\kappa_I = 0$, and $\gamma_I = 10.0$. Two different scenarios are considered, to concern both the hyperbolic and the parabolic limit of the system of equations. In the hyperbolic configuration, the relaxation times of all the compartments of individuals are $\tau = 1.0$, with the square of the characteristic velocities $\lambda^2 = 1.0$; while in the parabolic configuration $\tau = 10^{-5}$ and $\lambda^2 = 10^5$. The problem is solved applying an asymptotic-preserving (AP) Implicit-Explicit (IMEX) Runge–Kutta Finite Volume method, which permits to consistently simulate the diffusive (and stiff) regime of the system without loosing the expected 2^{nd} order accuracy [18, 22]. In Fig. 13, numerical results for both the scenarios are reported. A temporary persistence of the infectious can be noticed, with oscillations that reflect the sinusoidal form of the spatially variable contact rate. Differences of the dynamics of the epidemics in the two configurations of the relaxation times are evident. In particular, observing

the evolution of susceptible individuals, it can be seen that in the purely diffusive case the amount of susceptible tends to a much lower equilibrium value than in the hyperbolic case, with almost all the individuals of the system infected by the disease.

4.1.6 Application to the Emergence of COVID-19 in Italy

To analyze the effectiveness of the proposed approach in a realistic epidemic scenario, we design a numerical test reproducing the evolution of the first outbreak of COVID-19 in the Lombardy Region of Italy, from February 27, 2020 to March 27, 2020, with respect to uncertainties underlying the initial conditions and chosen epidemic parameters, considering the SEIAR-type multiscale transport SEIAR (83)–(84) in a network configuration, as described in Sect. 4.1.4. The system of equations is solved using a stochastic AP (sAP) IMEX Runge–Kutta Finite Volume Collocation method [19, 21, 123]. This numerical scheme permits to reach spectral accuracy in the stochastic space, if the solution is sufficiently smooth in that space, and to switch from a stochastic Collocation method for the advection problem to a stochastic Collocation method for the diffusive problem in a uniform way with respect to the involved parameters without loosing accuracy, i.e., sAP property [70, 71]. For further details regarding the numerical method and its convergence analysis, the reader can refer to [19].

A five-node network is considered, whose nodes represent the 5 main provinces interested by the epidemic outbreak in the first months of 2020: Lodi (n_1), Milan (n_2), Bergamo (n_3), Brescia (n_4), and Cremona (n_5). The arcs a_j connecting each node to the others identify the main set of routes and railways viable by commuters each day. A schematic representation of this network is shown in Fig. 14.

The transmission coefficients at each arc–node interface, as well as the percentage of commuters belonging to each province, are imposed using the official national mobility flow assessment. In particular, the matrix of commuters used reflects mobility data provided by Lombardy Region for the regional fluxes of year 2020 (see Data Sources in Sect. 5.1).

The characteristic speed associated to each arc is fixed to permit a full round trip in each origin–destination section within a day. The characteristic speed of compartment I is fixed to zero in all the nodes of the network. In the arcs, the relaxation time is assigned so that the model recovers a hyperbolic regime, while a parabolic setting is prescribed in the cities for commuters to simulate the diffusive behavior of the disease spread, which typically occurs in highly urbanized zones.

Concerning initial conditions and epidemic parameters of the test, at the beginning of the pandemic tracking of positive individuals cannot be considered reliable, but an information affected by uncertainty. To this aim, we introduce a single source of uncertainty z having uniform distribution, $z \sim \mathcal{U}(0, 1)$, and the initial conditions for compartment I, at each node, are prescribed as

$$I(x, 0, z) = I^0(1 + z)\,, \tag{87}$$

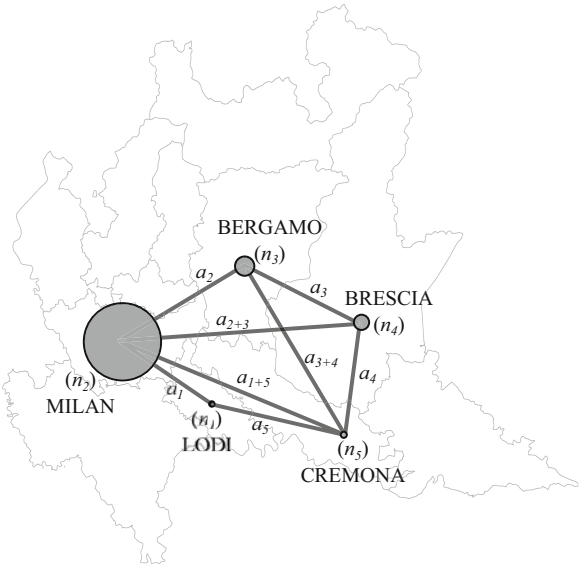


Fig. 14 Representation of the network of the Lombardy test case, composed of 5 nodes, corresponding to the provinces of interest and 5 arcs, connecting each city to the others, considering all the main paths of commuters. The dimension of the node is proportional to the dimension of the urbanized area of the province

with I^0 density of infectious people on February 27, 2020, as given by data recorded by the Civil Protection Department of Italy. The amount of total inhabitants of each province is given by 2019 data of the Italian National Institute of Statistics (see Data Sources in Sect. 5.1).

Due to the adopted screening policy, we chose to associate all infected individuals detected to the I compartment. Furthermore, also β_I is considered a random parameter:

$$\beta_I(0, z) = \beta_{I,0}(1 + \mu z) .$$

Assuming that highly infectious subjects are mostly detected in the most optimistic scenario, being subsequently quarantined or hospitalized, we set the minimum value $\beta_{I,0} = 0.03\,\beta_A$, as in [28, 57] and $\mu = 0.06^{-1}$. The initial value of β_A is calibrated as the result of a least square problem, namely the L2 norm of the difference between the observed cumulative number of infected $I(t)$ and the numerical evolution of the same compartment, through a deterministic SEIAR ODE model set up for the whole Lombardy Region, with the result $\beta_A = 0.545$. In the above fitting, we also estimated $E^0 \approx 10\,I^0$ and $A^0 \approx 9\,I^0$. Consequently, also initial conditions for compartments E, A, and S are stochastic, depending on the initial amount of severe infectious at each location, while $R^0 = 0$ everywhere. Finally, we fix γ_I, γ_A, and a according to [28, 57], considering these clinical parameters deterministic and σ as in [28, 74], setting then $\kappa_I = \kappa_A = 30$. With the above setup, we obtain an initial expected value of the basic reproduction number in the whole network $\mathbb{E}[R_0] = 3.6$, which is in agreement with estimations reported in [28, 57, 118].

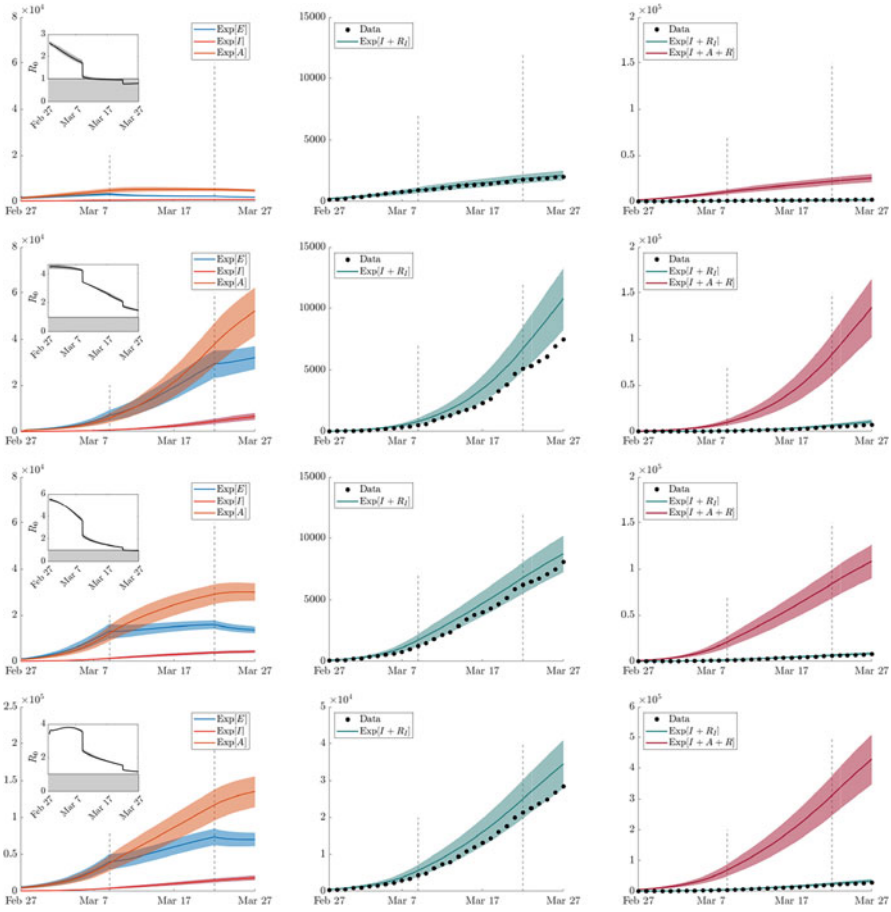


Fig. 15 Expected evolution in time, with 95% confidence intervals, for the cities of Lodi (first row), Milan (second row), Bergamo (third row), and the Lombardy network (fourth row), of: compartments E, A, I, together with the basic reproduction number R_0 (left); cumulative amount of severe infectious ($I + R_I$) compared with data of cumulative infectious taken from the Civil Protection Department of Italy (middle); cumulative amount of severe infectious ($I + R_I$) with respect to the effective cumulative amount of total infectious people, including asymptomatic individuals ($I + A + R$) (right). Vertical dashed lines identify the onset of governmental lockdown restrictions

We model the escalation of lockdown restrictions, starting from March 9, 2020, initial day of the northern Italy lockdown, reducing the transmission rates, increasing κ coefficients, due to the public being increasingly aware of the epidemic risks and reducing the percentage of commuting individuals according to mobility data tracked through mobile phones and made available by Google [1, 118].

Numerical results are reported in Fig. 15 for 3 representative cities, namely Lodi, Milan, Bergamo, and the whole Lombardy network. In Fig. 15 (first column) is

shown the expected evolution in time of the infected individuals, together with 95% confidence intervals, including E, I, and A. Each plot is also associated with the temporal evolution of the reproduction number $R_0(\mathbf{z}, t)$. One can see the capacity of the model to reproduce a very heterogeneous epidemic trend in the network analyzed, which is also reflected in the different ranges and patterns shown for the R_0 of each province. It can also be observed the agreement between the evolution of the reproduction number and the epidemic spread. In particular, it is confirmed the decline of the daily number of infected as R_0 reaches values below 1, as shown in the plots for Lodi and Bergamo. On the other hand, the persistence of the virus in the complete network, and especially in Milan, is noticed until March 27, 2020 (last day of the simulation), where the reproduction number remains greater than 1.

As visible from Fig. 15 (second column), the lower bound of the confidence band of the cumulative amount in time of I is comparable with the observed data of the Civil Protection Department of Italy. As expected, the mean value of infected people is higher, especially in Milan, the province most affected by the virus, due to the uncertainty of available data, which certainly underestimate the real amount of infected people.

The comparison between the expected evolution in time of the cumulative amount of I with respect to the effective cumulative amount of total infectious people, including also compartment A, is shown in Fig. 15 (third column). Here, it can be noticed how much of the spread of COVID-19 has actually been lost from the data of the first outbreak in Lombardy and the impact that the presence of asymptomatic or undetected subjects has had on the epidemic evolution.

4.2 Realistic Geographical Settings

4.2.1 2D Kinetic Transport Model

Let us now define $\Omega \subset \mathbb{R}^2$ a two-dimensional geographical area of interest, still assuming that individuals have been separated into commuting and non-commuting population, with the former at position $x \in \Omega$ moving with velocity directions $v \in \mathbb{S}^1$. Considering initially a simple SIR-dynamics, we denote by $f_S = f_S(\mathbf{z}, x, v, t)$, $f_I = f_I(\mathbf{z}, x, v, t)$, and $f_R = f_R(\mathbf{z}, x, v, t)$, where $\mathbf{z} \in \mathbb{R}^d$ is a random vector characterizing the possible sources of uncertainty, the respective kinetic densities of susceptible, infected, and recovered individuals. The kinetic distribution of commuters is then given by:

$$f(\mathbf{z}, x, v, t) = f_S(\mathbf{z}, x, v, t) + f_I(\mathbf{z}, x, v, t) + f_R(\mathbf{z}, x, v, t),$$

and their total density is obtained by integration over the velocity space

$$\rho(\mathbf{z}, x, t) = \frac{1}{2\pi} \int_{\mathbb{S}^1} f(\mathbf{z}, x, v_*, t)\, dv_*.$$

As a consequence, the number of commuting susceptible, infectious, and recovered individuals can be recovered irrespective of their direction of displacement by integration over the velocity space. This gives

$$S_c(\mathbf{z}, x, t) = \frac{1}{2\pi} \int_{\mathbb{S}^1} f_S(\mathbf{z}, x, v, t)\, dv\,, \quad I_c(\mathbf{z}, x, t) = \frac{1}{2\pi} \int_{\mathbb{S}^1} f_I(\mathbf{z}, x, v, t)\, dv\,,$$

$$R_c(\mathbf{z}, x, t) = \frac{1}{2\pi} \int_{\mathbb{S}^1} f_R(\mathbf{z}, x, v, t)\, dv\,.$$

In this setting, the densities of the commuters satisfy the kinetic transport equations [23]

$$\begin{aligned}
&\frac{\partial f_S}{\partial t} + \nabla_x \cdot (v_S f_S) = -F_I(f_S, I) + \frac{1}{\tau_S}(S_c - f_S)\,,\\
&\frac{\partial f_I}{\partial t} + \nabla_x \cdot (v_I f_I) = F_I(f_S, I) - \gamma_I f_I + \frac{1}{\tau_I}(I_c - f_I)\,,\\
&\frac{\partial f_R}{\partial t} + \nabla_x \cdot (v_R f_R) = \gamma_I f_I + \frac{1}{\tau_R}(R_c - f_R)\,,
\end{aligned} \tag{88}$$

where the total densities are still defined by the sum of commuting and non-commuting part $S = S_c + S_0$, $I = I_c + I_0$, and $R = R_c + R_0$. Densities of non-commuters, who act only at a local scale, satisfy the following diffusion dynamics:

$$\begin{aligned}
&\frac{\partial S_0}{\partial t} = -F_I(S_0, I) + \nabla_x \cdot (D_S \nabla_x S_0),\\
&\frac{\partial I_0}{\partial t} = F_I(S_0, I) - \gamma_I I_0 + \nabla_x \cdot (D_I \nabla_x I_0),\\
&\frac{\partial R_0}{\partial t} = \gamma_I I_0 + \nabla_x \cdot (D_R \nabla_x R_0).
\end{aligned} \tag{89}$$

In the resulting multiscale kinetic SIR model (88)–(89), which couples the commuting and non-commuting dynamics, the velocities $v_i = \lambda_i v$ in (88), as well as the diffusion coefficients D_i in (89), with $i \in \{S, I, R\}$, are designed to take into account the heterogeneity of geographical areas and are thus chosen dependent

on the spatial location. The same stands also for the relaxation times, in analogy with system (76)–(77). We refer to (78) for the definition of the incidence function $F_I(\cdot, I)$, and to Sect. 4.1.1 in general for the definition of the epidemic parameters involved. Furthermore, for the definition of the reproduction number of the above system we refer to (82) and the relative discussion.

4.2.2 Macroscopic Formulation and Diffusion Limit

Let us introduce the flux functions

$$J_S = \frac{\lambda_S}{2\pi}\int_{\mathbb{S}^1} v f_S \, dv, \quad J_I = \frac{\lambda_I}{2\pi}\int_{\mathbb{S}^1} v f_I \, dv, \quad J_R = \frac{\lambda_R}{2\pi}\int_{\mathbb{S}^1} v f_R \, dv, \tag{90}$$

where we omitted the dependence on (x, v, t) for notation simplicity. Assuming λ_i, $i \in \{S, I, R\}$, independent from space and integrating system (88) in v, we get the following set of equations for the macroscopic formulation of densities:

$$\begin{aligned} &\frac{\partial S_c}{\partial t} + \nabla_x \cdot J_S = -F_I(S_c, I), \\ &\frac{\partial I_c}{\partial t} + \nabla_x \cdot J_I = F_I(S_c, I) - \gamma_I I_c, \\ &\frac{\partial R_c}{\partial t} + \nabla_x \cdot J_R = \gamma_I I_c, \end{aligned} \tag{91}$$

and fluxes

$$\begin{aligned} &\frac{\partial J_S}{\partial t} + \frac{\lambda_S^2}{2\pi}\int_{\mathbb{S}^1} (v \cdot \nabla_x f_S) v \, dv = -F_I(J_S, I) - \frac{1}{\tau_S} J_S, \\ &\frac{\partial J_I}{\partial t} + \frac{\lambda_I^2}{2\pi}\int_{\mathbb{S}^1} (v \cdot \nabla_x f_I) v \, dv = \frac{\lambda_I}{\lambda_S} F_I(J_S, I) - \gamma_I J_I - \frac{1}{\tau_I} J_I, \\ &\frac{\partial J_R}{\partial t} + \frac{\lambda_R^2}{2\pi}\int_{\mathbb{S}^1} (v \cdot \nabla_x f_R) v \, dv = \frac{\lambda_R}{\lambda_I}\gamma_I J_I - \frac{1}{\tau_R} J_R. \end{aligned} \tag{92}$$

Note that the above system is not closed because the evolution of the fluxes in (92) involves higher order moments of the kinetic densities.

The diffusion limit can be formally recovered by introducing the diffusion coefficients $D_i = \frac{1}{2}\lambda_i^2 \tau_i$, with $i \in \{S, I, R\}$, and letting $\tau_i \to 0$. We get, from the r.h.s. in (88), $f_S = S_c$, $f_I = I_c$, $f_R = R_c$, and, consequently, from (92) we recover Fick's laws

$$J_S = -D_S \nabla_x S_c, \quad J_I = -D_I \nabla_x I_c, \quad J_R = -D_R \nabla_x R_c,$$

which inserted into (91) lead to the diffusion system for the population of commuters [86, 110, 121]

$$\begin{aligned}\frac{\partial S_c}{\partial t} &= -F_I(S_c, I) + \nabla_x \cdot (D_S \nabla_x S_c)\\ \frac{\partial I_c}{\partial t} &= F_I(S_c, I) - \gamma_I I_c + \nabla_x \cdot (D_I \nabla_x I_c)\\ \frac{\partial R_c}{\partial t} &= \gamma_I I_c + \nabla_x \cdot (D_R \nabla_x R_c)\end{aligned} \tag{93}$$

coupled with (89) for the non-commuting counterpart.

Similarly to the one-dimensional case, the capability of the model to account for different regimes, hyperbolic or parabolic, according to the space dependent relaxation times τ_i, $i \in \{S, I, R\}$, makes it suitable for describing the dynamics of human beings. Indeed, it is reasonable to avoid describing the details of movements within an urban area and model this through a diffusion operator. On the other hand, commuters when moving from one city to another follow well-established connections for which a description via transport operators is more appropriate.

4.2.3 Extension to Multi-Compartmental Modelling

As previously presented for the 1D model in Sect. 4.1.3, it is possible to extend the modelling considering more general compartmental subdivisions. For example, more realistic models for COVID-19 should take into account the exposed population as well as the asymptomatic fraction of infected. As an example, we describe the extension of the multiscale kinetic transport modelling presented in the previous sections to a more general compartmental structure, where the exposed population is included and infected people are distinguished between highly symptomatic and mildly/no symptomatic (see Fig. 11). We denote the commuter individuals who belong to the newly introduced compartment of exposed by $f_E(\mathbf{z}, x, v, t)$ and of asymptomatic (or mildly symptomatic) by $f_A(\mathbf{z}, x, v, t)$, then the total density results

$$f(\mathbf{z}, x, v, t) = \sum_{i \in \{S,E,I,A,R\}} f_i(\mathbf{z}, x, v, t).$$

The kinetic SEIAR-type dynamics of the commuters then reads [20]

$$
\begin{aligned}
&\frac{\partial f_S}{\partial t} + \nabla_x \cdot (v_S f_S) = -F_I(f_S, I) - F_A(f_S, A) + \frac{1}{\tau_S}(S_c - f_S)\,,\\
&\frac{\partial f_E}{\partial t} + \nabla_x \cdot (v_E f_E) = F_I(f_S, I) + F_A(f_S, A) - a f_E + \frac{1}{\tau_E}(E_c - f_E)\,,\\
&\frac{\partial f_I}{\partial t} + \nabla_x \cdot (v_I f_I) = a\sigma f_E - \gamma_I f_I + \frac{1}{\tau_I}(I_c - f_I)\,,\\
&\frac{\partial f_A}{\partial t} + \nabla_x \cdot (v_A f_A) = a(1-\sigma) f_E - \gamma_A f_A + \frac{1}{\tau_A}(A_c - f_A)\,,\\
&\frac{\partial f_R}{\partial t} + \nabla_x \cdot (v_R f_R) = \gamma_I f_I + \gamma_A f_A + \frac{1}{\tau_R}(R_c - f_R)\,,
\end{aligned}
\tag{94}
$$

with $E = E_c + E_0$, $A = A_c + A_0$, and

$$
E_c(\mathbf{z}, x, t) = \frac{1}{2\pi}\int_{\mathbb{S}^1} f_E(\mathbf{z}, x, v, t)\, dv\,, \quad A_c(\mathbf{z}, x, t) = \frac{1}{2\pi}\int_{\mathbb{S}^1} f_A(\mathbf{z}, x, v, t)\, dv\,.
$$

Indeed, this system is coupled with the following one describing the dynamics of non-commuters, who act only at the urban scale:

$$
\begin{aligned}
\frac{\partial S_0}{\partial t} &= -F_I(S_0, I) - F_A(S_0, A) + \nabla_x \cdot (D_S \nabla_x S_0),\\
\frac{\partial E_0}{\partial t} &= F_I(S_0, I) + F_A(S_0, A) - a E_0 + \nabla_x \cdot (D_E \nabla_x E_0),\\
\frac{\partial I_0}{\partial t} &= a\sigma E_0 - \gamma_I I_0 + \nabla_x \cdot (D_I \nabla_x I_0),\\
\frac{\partial A_0}{\partial t} &= a(1-\sigma) E_0 - \gamma_A A_0 + \nabla_x \cdot (D_A \nabla_x A_0),\\
\frac{\partial R_0}{\partial t} &= \gamma_I I_0 + \gamma_A A_0 + \nabla_x \cdot (D_R \nabla_x R_0).
\end{aligned}
\tag{95}
$$

For the definition of the incidence function regarding asymptomatic people $F_A(\cdot, A)$, we consider (85).

When introducing the same definition of flux (90) for the additional compartments J_E and J_A, and integrating system (94) in v, we get the set of equations for the macroscopic densities. Moreover, defining also $D_E = \frac{1}{2}\lambda_E^2 \tau_E$ and $D_A = \frac{1}{2}\lambda_A^2 \tau_A$, assuming the velocities λ_i, $i \in \{S, E, I, A, R\}$, independent from space, and considering the analogous of the diffusion limit discussed in Sect. 4.2.2, we recover SEIAR-type system in the diffusive regime for the commuting individuals [20] coupled with (95) for the non-commuting counterpart.

To define the reproduction number also for this multiscale SEIAR-type kinetic transport model, we recall the NGM approach [41] considering no flux boundary

conditions, which yields the same definition (86). Details of this derivation are reported in [20].

4.2.4 Application to the Spatial Spread of COVID-19 in Italy in Emilia-Romagna and Lombardy Region

Let us underline that the discretization of the resulting multiscale systems of PDEs is not trivial and therefore requires the construction of a specific numerical method able to correctly describe the transition from a convective to a diffusive regime in realistic geometries. For this purpose, we adopt an asymptotic-preserving IMEX Runge–Kutta method on unstructured grids coupled with a stochastic Collocation method that ensures spectral accuracy in the stochastic space [68, 71, 94]. At each collocation node, the numerical scheme combines a discrete ordinate method in velocity with the even and odd parity formulation [42, 69] and achieves asymptotic preservation in time using suitable IMEX Runge–Kutta schemes [22], namely, to obtain a scheme that consistently captures the diffusion limit and for which the choice of the time discretization step is not related to the smallness of the scaling parameters τ. All the details concerning the numerical scheme and its validation in terms of accuracy are reported in [20, 23].

To validate the proposed methodology in realistic geographical and epidemiological scenarios, two numerical tests reproducing, respectively, the epidemic outbreak of COVID-19 in the Emilia-Romagna Region of Italy, from March 1, 2020 to March 10, 2020, and in the Lombardy Region of Italy, from February 27, 2020 to March 22, 2020, are designed. In the former, we solve a multiscale SEIR-type system of PDEs in a deterministic setting (for further details on the chosen SEIR model the reader can refer to [23]). In the latter, we also take into account the uncertainty underlying initial conditions of infected individuals, solving the multiscale SEIAR-type system of PDEs (88)–(89).

The computational domain is defined in terms of the boundary that circumscribes the Regions as a list of geo-referenced points in the ED50/UTM Zone 32N reference coordinate system from Istituto Nazionale di Statistica (see Data Sources in Sect. 5.1). No-flux boundary conditions are imposed in the whole boundary of the domain, assuming that the population is not moving from/to the adjacent Regions. The domain is then subdivided in the provinces of the specific Region. The identification of these cities is shown in Fig. 16 (top left) for Emilia-Romagna and in Fig. 16 (bottom left) for Lombardy. To avoid the mobility of the population in the entire territory and to simulate a more realistic geographical scenarios in which individuals travel along the main traffic paths of the Region, different values of propagation speeds are assigned in the domain, which reflect, as close as possible, the real characteristics of the territory. The resulting distribution of the characteristic speeds is visible from Fig. 16 (top right) for Emilia-Romagna and Fig. 16 (bottom right) for Lombardy case.

The space-dependent relaxation time is assigned so that the model recovers a hyperbolic regime in the entire region, apart from the main cities, where a parabolic

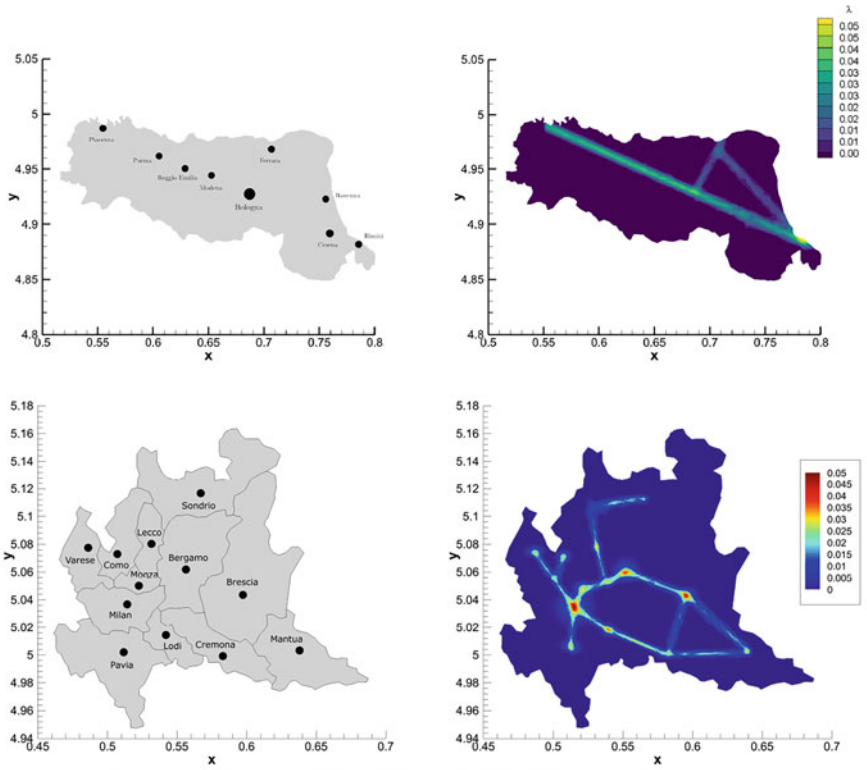


Fig. 16 Top: identification of the Emilia-Romagna provinces (left) and initial condition imposed for characteristic speeds λ_i, $i \in \{S, E, R\}$ (right). Bottom: identification of the Lombardy provinces (left), initial condition imposed for characteristic speeds λ_i, $i \in \{S, E, A, R\}$ (right)

setting is prescribed to correctly capture the diffusive behavior of the disease spreading, which typically occurs in highly urbanized zones. Considering p_c the number of citizens of a generic city (province) denoted with subscript $c = 1, \ldots, \mathcal{N}_c$, the initial spatial distribution of the generic population $f(x, y)$ is assigned, for each province and each epidemiological compartment, as a multivariate Gaussian function with the variance being the radius of the urban area r_c:

$$f(x, y) = \frac{1}{2\pi r_c} e^{-\frac{(x-x_c)^2+(y-y_c)^2}{2r_c^2}} p_c \,,$$

with (x_c, y_c) representing the coordinates of a generic city center. The initial population setting, for each province, is taken from 2019 data of the Italian National Institute of Statistics.

For the Emilia-Romagna Region, we estimate the initial number of exposed individuals, including asymptomatic, as $E^0 = 4\, I^0$; while, for the Lombardy

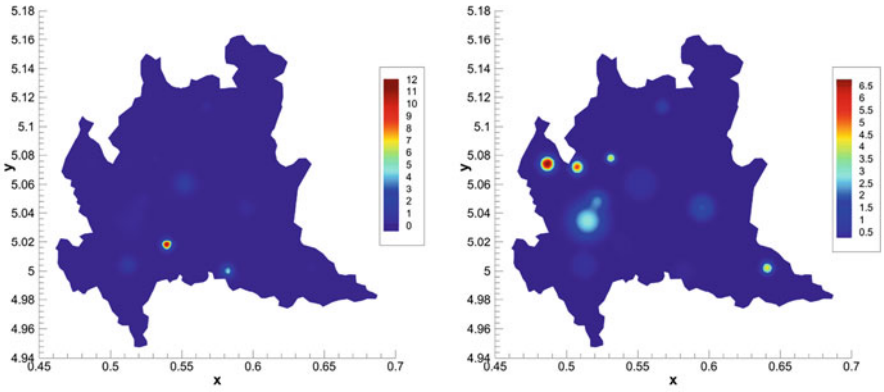


Fig. 17 Initial distribution (on February 27, 2020) of the infected population $E^0 + I^0 + A^0$ (left) and of the reproduction number $R_0(0)$ (right) in the Lombardy Region

Region, $E^0 = 10\, I^0$ and $A^0 = 9\, I^0$ in each location, with I^0 given by data recorded by the Civil Protection Department of Italy in the first day simulated. As previously discussed in Sect. 4.1.6, for the Lombardy test, we introduce a single source of uncertainty z having uniform distribution, $z \sim \mathcal{U}(0, 1)$ so that the initial conditions for compartment I, in each control volume, are prescribed as in (87). Moreover, we refer to regional mobility data to properly subdivide the population in commuters and non-commuters (see Data Sources in Sect. 5.1). Concerning the calibration and the choice of clinical epidemic parameters, as well as for the modelling of the governmental restrictions, the reader can refer to [23] for the Emilia-Romagna case and to [19] for the Lombardy case. With the chosen parametric setups, we obtain initial reproduction number for Emilia-Romagna $R_0 = 2.3$ and an initial expected value of the basic reproduction number for the Lombardy Region $\mathbb{E}[R_0(0)] = 3.2$, which are in accordance with the available literature [28, 57, 118]. Nevertheless, with the proposed methodology it is possible to present the heterogeneity underlying the basic reproduction number at the local scale, as shown for the Lombardy case in Fig. 17 (bottom), together with the initial global amount of infected people $E^0(x, y) + I^0(x, y) + A^0(x, y)$ present in the domain.

Figure 18 plots a comparison against the measured data at the Regional level, reported by the Civil Protection Department of Italy, and the same comparison for the province of Piacenza, Parma, Bologna, and Rimini, depicting an overall very good agreement. Figure 19 shows the time evolution of the exposed population E, including asymptomatic, which is moving from both Piacenza and Rimini toward the center of the region and the city of Bologna, then spreading northern in the direction of Ferrara. The wave of the exposed population is clearly visible, highlighting the hyperbolic regime of the model.

In Fig. 20 (first column), the expected evolution in time of the infected individuals, together with 95% confidence intervals, is shown for exposed E, highly symptomatic subjects I, and asymptomatic or weakly symptomatic people A,

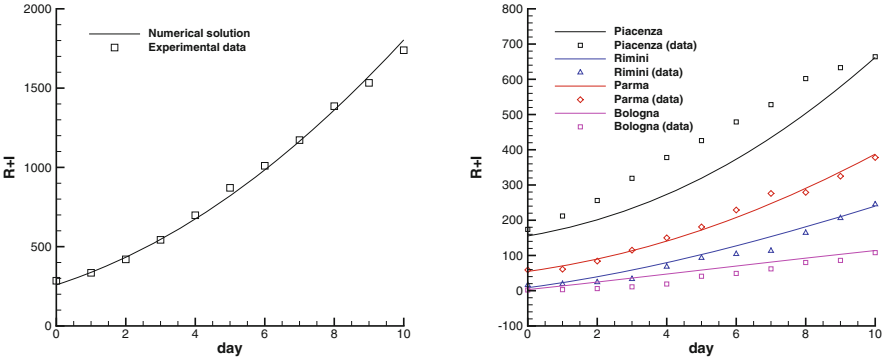


Fig. 18 Left: time evolution of total infected and recovered population ($R + I$) compared against experimental data for the Emilia-Romagna Region. Right: time evolution of total infected and recovered population ($R + I$) compared against experimental data for the province of Piacenza (black), Parma (red), Bologna (purple), and Rimini (blue)

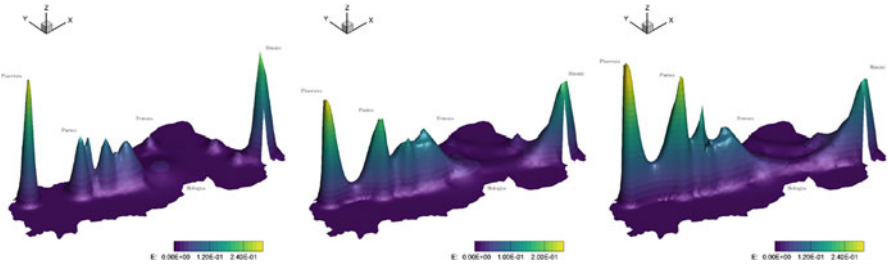


Fig. 19 Distribution of exposed population E, including asymptomatic, on March 1, 4, and 10, 2020 (from left to right) in Emilia-Romagna

for three representative cities, namely Lodi, Milan, and Bergamo, and the whole Lombardy Region. Here it is already appreciable the heterogeneity of the diffusion of the virus. Indeed, from the different y-axis scales adopted for the plot of the provinces, it can be noticed that Milan and Bergamo present a consistently higher contagion with respect to the one shown in Lodi. From the same Fig. 20 (second column), it can be observed that the lower bound of the confidence interval of the cumulative amount in time of highly symptomatic individuals is in line with data reported by the Civil Protection Department of Italy. As expected, due to the uncertainty taken into account, the mean value of the numerical result in each city is higher than the registered one. Also the comparison between the expected evolution in time of the cumulative amount of severe infectious with respect to the effective cumulative amount of total infectious people, including asymptomatic and mildly symptomatic individuals, is shown in Fig. 20 (third column). From this figure, it is clear that the number of infections recorded during the first outbreak of COVID-19 in Lombardy represents a clear underestimation of the actual trend of infection suffered by the Region and by Italy as a whole, and how the presence of

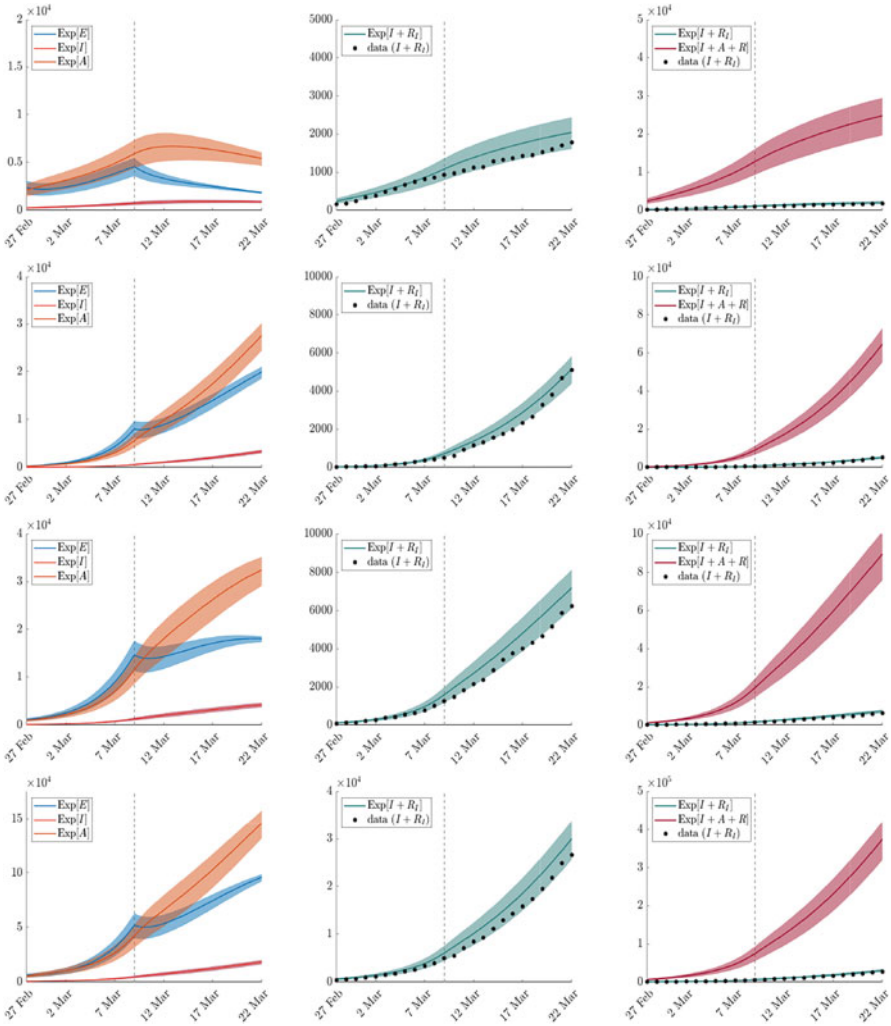


Fig. 20 Expected evolution in time, with 95% confidence intervals, for the cities of Lodi (first row), Milan (second row), Bergamo (third row), and the Lombardy region (fourth row). Compartments E, A, I (left); severe infectious ($I + R_I$) compared with data of cumulative infectious taken from the Civil Protection Department of Italy (middle); severe infectious ($I + R_I$) with respect to the effective cumulative amount of total infectious people, including asymptomatic individuals ($I + A + R$) (right). The vertical dashed line identifies governmental lockdown restrictions

asymptomatic subjects, not detected, has affected the pandemic evolution. Results concerning the rest of the cities of the Region can be found in [20].

In Fig. 21, final expectation and variance of the cumulative amount of infected people, namely $E + A + I$, are reported in the 2D framework of Lombardy. If

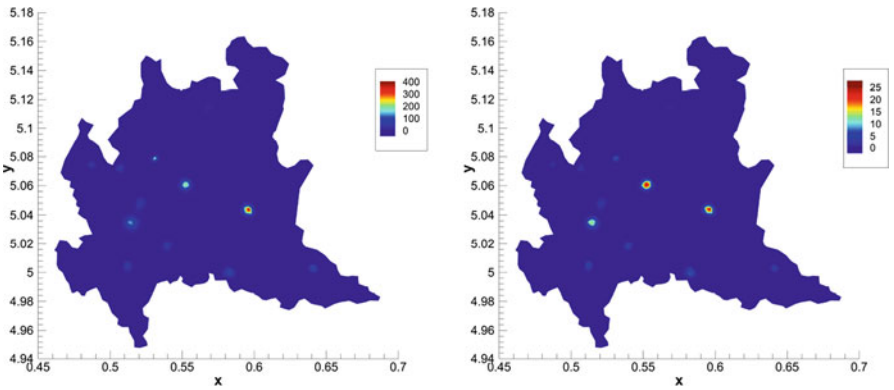


Fig. 21 Expectation (left) and variance (right) of the cumulative amount of infected people $E + A + I$ at the end of the simulation (March 22, 2020) in the Lombardy Region

comparing Fig. 21 (top left) with 16 (bottom left), it can be noticed that, at the end of March, the virus is no longer mostly affecting the province of Lodi and Cremona, but has been spread, arriving to hit most of all Brescia, Milan, and Bergamo.

5 Concluding Remarks and Research Perspectives

In this survey, we presented a series of recent results obtained in the field of kinetic modelling applied to epidemiology. In particular, we focused on three main aspects: the influence of social features such as the number of contacts, wealth and age of individuals, the design of effective control techniques even in the presence of uncertain data, and finally, the impact on the pandemic of the movements of individuals both on urban and extra-urban scales. All these aspects proved essential in order to present realistic scenarios on the spread of the epidemic and in agreement with the observed data.

The modelling approach presented here, although in some cases developed for the sake of simplicity on compartmental models with a very basic structure, can be easily extended, as analyzed in the last part of the survey, also to more realistic models for the spread of the COVID-19 epidemic. In particular, given the generality of the social dynamics modelling introduced here, this opens interesting perspectives in future directions by going to evaluate the impact of additional features that can influence the evolution of the pandemic, such as the viral load [38, 84] or the spread of fake news [52]. The former in particular plays a decisive role in analyzing the influence of the so-called super-spreaders [77, 92], while the latter as we have seen play a key role regarding the vaccination campaign [83, 103].

5.1 Data sources

With respect to the numerical results presented in the simulations, specifically in Sections 2.15, 3.2.2, 4.1.6 and 4.2.4, the following data repositories were used. The GitHub repository of the Italian Civil Protection Department:[1] the John Hopkins University GitHub repository [47]; Regione Lombardia, Italy, Commuters Data;[2] the Italian National Institute of Statistics, ISTAT;[3] Geographical Data from ISTAT;[4] and Regione Emilia-Romagna, Italy, Commuters Data.[5]

Acknowledgments This work has been written within the activities of GNFM and GNCS groups of INdAM (National Institute of High Mathematics). G.A., G.B., W.B., G.D., and L. P. acknowledge the support of MIUR-PRIN Project 2017, No. 2017KKJP4X *Innovative numerical methods for evolutionary partial differential equations and applications*. GA also acknowledges partial support from the Program Ricerca di Base 2019 of the University of Verona entitled "Geometric Evolution of Multi-Agent Systems." GB holds a Research Fellowship from INdAM. G.T. and M. Z. were partially supported by the MIUR Program (2018–2022), *Dipartimenti di Eccellenza*, Department of Mathematics, University of Pavia.

References

1. A. Aktay, et al., Google COVID-19 community mobility reports: anonymization process description (version 1.1) (2020). arXiv:2004.04145
2. G. Albi, L. Pareschi, Selective model-predictive control for flocking systems. Commun. Appl. Ind. Math. **9**(2), 4–21 (2018)
3. G. Albi, L. Pareschi, M. Zanella, Boltzmann-type control of opinion consensus through leaders. Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. **372**(2028), 20140138 (2014)
4. G. Albi, M. Herty, L. Pareschi, Kinetic description of optimal control problems and applications to opinion consensus. Commun. Math. Sci. **13**(6), 1407–1429 (2015)
5. G. Albi, L. Pareschi, M. Zanella, Uncertainty quantification in control problems for flocking models. Math. Probl. Eng. **2015**, 850124 (2015)
6. G. Albi, L. Pareschi, M. Zanella, Opinion dynamics over complex networks: kinetic modeling and numerical methods. Kinet. Relat. Models **10**(1), 1–32 (2017)
7. G. Albi, N. Bellomo, L. Fermo, S.Y. Ha, J. Kim, L. Pareschi, D. Poyato, J. Soler, Vehicular traffic, crowds, and swarms: from kinetic theory and multiscale methods to applications and research perspectives. Math. Models Methods Appl. Sci. **29**(10), 1901–2005 (2019)
8. G. Albi, L. Pareschi, M. Zanella, Control with uncertain data of socially structured compartmental models. J. Math. Biol. **82**, 63 (2021)

[1] https://github.com/pcm-dpc/COVID-19.

[2] https://www.dati.lombardia.it/Mobilit-e-trasporti/Matrice-OD2020-Passeggeri/hyqr-mpe2.

[3] https://demo.istat.it/.

[4] https://www4.istat.it/it/archivio/209722.

[5] https://sasweb.regione.emilia-romagna.it/statistica/SceltaAnno.do?analisi=matPend2011_2015.

9. G. Albi, L. Pareschi, M. Zanella, Modelling lockdown measures in epidemic outbreaks using selective socio-economic containment with uncertainty. Math. Biosci. Eng. **18**(6), 7161–7190 (2021)
10. R.M. Anderson, R.M. May, *Infectious Diseases of Humans: Dynamics and Control* (Oxford University Press, Oxford, 1991)
11. B. Aylaj, N. Bellomo, L. Gibelli, A. Reali, A unified multiscale vision of behavioral crowds. Math. Mod. Methods Appl. Sci. **30**(1), 1–22 (2020)
12. E. Barbera, G. Consolo, G. Valenti, Spread of infectious diseases in a hyperbolic reaction-diffusion susceptible-infected-recovered model. Phys. Rev. E **88**, 052719 (2013)
13. B. Barthélemy, A. Barrat, R. Pastor-Satorras, A. Vespignani, Dynamical patterns of epidemic outbreaks in complex heterogeneous networks. J. Theor. Biol. **235**, 275–288 (2005)
14. N. Bellomo, L. Gibelli, N. Outada, On the interplay between behavioral dynamics and social interactions in human crowds. Kinet. Relat. Models **12**(2), 397–409 (2019)
15. N. Bellomo, R. Bingham, M. A. J. Chaplain, G. Dosi, G. Forni, D.A. Knopoff, J. Lowengrub, R. Twarock, M.E. Virgillito, A multiscale model of virus pandemic: heterogeneous interactive entities in a globally connected world. Math. Models Methods Appl. Sci. **30**(8), 1591–1651 (2020)
16. G. Béraud, et al., The French connection: the first large population-based contact survey in france relevant for the spread of infectious diseases. PLoS ONE **10**(7), e0133203 (2015)
17. H. Berestycki, J.M. Roquejoffre, L. Rossi, Propagation of epidemics along lines with fast diffusion. Bull. Math. Biol. **83**, 2 (2021)
18. G. Bertaglia, L. Pareschi, Hyperbolic models for the spread of epidemics on networks: kinetic description and numerical methods. ESAIM: Math. Model. Numer. Anal. **55**, 381–407 (2021)
19. G. Bertaglia, L. Pareschi, Hyperbolic compartmental models for epidemic spread on networks with uncertain data: application to the emergence of Covid-19 in Italy. Math. Models Methods Appl. Sci. **31**(12), 2495–2531 (2021)
20. G. Bertaglia, W. Boscheri, G. Dimarco, L. Pareschi, Spatial spread of COVID-19 outbreak in Italy using multiscale kinetic transport equations with uncertainty. Math. Biosci. Eng. **18**(5), 7028–7059 (2021)
21. G. Bertaglia, V. Caleffi, L. Pareschi, A. Valiani, Uncertainty quantification of viscoelastic parameters in arterial hemodynamics with the a-FSI blood flow model. J. Comput. Phys. **430**, 110102 (2021)
22. S. Boscarino, L. Pareschi, G. Russo, A unified IMEX Runge-Kutta approach for hyperbolic systems with multiscale relaxation. SIAM J. Numer. Anal. **55**(4), 2085–2109 (2017)
23. W. Boscheri, G. Dimarco, L. Pareschi, Modeling and simulating the spatial spread of an epidemic through multiscale kinetic transport equations. Math. Model. Methods Appl. Sci. **31**(6), 1059–1097 (2021)
24. J.F. Bouchaus, Mézard: wealth condensation in a simple model of economy. Phys. A **282**, 536 (2000)
25. F. Brauer, C. Castillo-Chavez, Z. Feng, *Mathematical Models in Epidemiology*. Texts in Applied Mathematics, vol. 69 (Springer, New York, 2019)
26. G. Bretti, R. Natalini, M. Ribot, A hyperbolic model of chemotaxis on a network: a numerical study. ESAIM: Math. Model. Numer. Anal. **48**(1), 231–258 (2014)
27. T. Britton, F. Ball, P. Trapman, A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2. Science **369**, 6505 (2020)
28. B. Buonomo, R. Della Marca, Effects of information-induced behavioural changes during the COVID-19 lockdowns: the case of Italy: COVID-19 lockdowns and behavioral change. R. Soc. Open Sci. **7**, 201635 (2020)
29. A. Capaldi, et al., Parameter estimation and uncertainty quantification for an epidemic model. Math. Biosci. Eng. **9**(3), 553–576 (2012)
30. V. Capasso, G. Serio, A generalization of the Kermack-McKendrick deterministic epidemic model. Math. Biosci. **42**, 43–61 (1978)

31. C. Castillo-Chavez, H.W. Hethcote, V.A. Andreasen, S.A. Levin, W.M. Liu, Epidemiological models with age structure, proportionate mixing, and cross-immunity. J. Math. Biol. **27**(3), 233–258 (1989)
32. C. Cercignani, *The Boltzmann Equation and its Applications*. Springer Series in Applied Mathematical Sciences, vol. 67 (Springer, New York, 1988)
33. G. Chowell, Fitting dynamic models to epidemic outbreaks with quantified uncertainty: a primer for parameter uncertainty, identifiability, and forecast. Infect. Dis. Model. **2**(3), 379–398 (2017)
34. A. Ciallella, M. Pulvirenti, S. Simonella, Kinetic SIR equations and particle limits. Atti Accad. Naz. Lincei Rend. Lincei Mat. Appl. **32**(2), 295–315 (2021)
35. R.M. Colombo, M. Garavello, F. Marcellini, E. Rossi, An age and space structured SIR model describing the Covid-19 pandemic. J. Math. Ind. **10**, 22 (2020)
36. S. Cordier, L. Pareschi, G. Toscani, On a kinetic model for a simple market economy. J. Stat. Phys. **120**, 253–277 (2005)
37. S. Cordier, D. Maldarella, L. Pareschi, C. Piatecki, Microscopic and kinetic models in financial markets, in *Mathematical Modeling of Collective Behavior in Socio-Economic and Life Sciences*, ed. by G. Naldi, L. Pareschi, G. Toscani. Modeling and Simulation in Science, Engineering and Technology (Birkhäuser Boston, 2010), pp. 51–80
38. R. Della Marca, N. Loy, A. Tosin, An SIR-like kinetic model tracking individuals' viral load 2021 (2021, preprint). arXiv:2106.14480
39. F. Della Rossa, et al., A network model of Italy shows that intermittent regional strategies can alleviate the COVID-19 epidemic. Nat. Commun. **11**(1), 1–9 (2020)
40. O. Diekmann, J.A.P. Heesterbeek, *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation* (Wiley, Chichester, 2000)
41. O. Diekmann, J.A.P. Heesterbeek, J.A.J. Metz, On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. J. Math. Biol. **28**(4), 365–382 (1990)
42. G. Dimarco, L. Pareschi, Numerical methods for kinetic equations. Acta Numer. **23**, 369–520 (2014)
43. G. Dimarco, G. Toscani, Kinetic modeling of alcohol consumption. J. Stat. Phys. **177**, 1022–1042 (2019)
44. G. Dimarco, L. Pareschi, G. Toscani, M. Zanella, Wealth distribution under the spread of infectious diseases. Phys. Rev. E **102**, 022303 (2020)
45. G. Dimarco, B. Perthame, G. Toscani, M. Zanella, Kinetic models for epidemic dynamics with social heterogeneity. J. Math. Biol. **83**, 4 (2021)
46. G. Dimarco, G. Toscani, M. Zanella, Optimal control of epidemic spreading in presence of social heterogeneity (2021, preprint). arXiv:2107.12180
47. E. Dong, H. Du, L. Gardner, An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect. Dis. **20**(5), 533–534 (2020)
48. B. Düring, L. Pareschi, G. Toscani, Kinetic models for optimal control of wealth inequalities. Eur. Phys. J. B **91**, 265 (2018)
49. R. Dutta, S. Gomes, D. Kalise, L. Pacchiardi, Using mobility data in the design of optimal lockdown strategies for the COVID-19 pandemic. PLoS Comput. Biol. **17**(8), e1009236 (2020)
50. M.S. Eichenbaumz, S. Rebelox, M. Trabandt, The macroeconomics of epidemics. NBER Working Paper 26882, National Bureau of Economic Research (2020)
51. S. Flaxman, et al., Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries. Report 13. Imperial College COVID-19 Response Team (2020)
52. J. Franceschi, L. Pareschi, On the interplay between fake news dissemination and competence: kinetic description and numerical approximation (preprint, 2021). arXiv:2109.14087
53. E. Franco, A feedback SIR (fSIR) model highlights advantages and limitations of infection-based social distancing (2020, preprint). arXiv:2004.13216v3

54. K.O. Friedrichs, P.D. Lax, Systems of conservation equations with a convex extension. Proc. Natl. Acad. Sci. **68**(8), 1686–1688 (1971)
55. G. Furioli, A. Pulvirenti, E. Terraneo, G. Toscani, Fokker–Planck equations in the modelling of socio-economic phenomena. Math. Models Methods Appl. Sci. **27**(1), 115–158 (2017)
56. G. Furioli, A. Pulvirenti, E. Terraneo, G. Toscani, Non-Maxwellian kinetic equations modeling the evolution of wealth distribution. Math. Models Methods Appl. Sci. **30**(4), 685–725 (2020)
57. M. Gatto, et al., Spread and dynamics of the COVID-19 epidemic in Italy: effects of emergency containment measures. Proc Natl Acad Sci U. S. A. **117**(19), 10484–10491 (2020)
58. A. Ghosh, A. Chatterjee, J.I. Inoue, B.K. Chakrabarti, Inequality measures in kinetic exchange models of wealth distributions. Phys. A **451**, 465 (2016)
59. G. Giordano, et al., Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nat. Med. **26**, 855–860 (2020)
60. J. Glasser, Z. Feng, A. Moylan, S. Del Valle, C. Castillo-Chavez, Mixing in age-structured population models of infectious diseases. Math. Bios. **235**(1), 1–7 (2012)
61. S. Gualandi, G. Toscani, Human behavior and lognormal distribution. A kinetic description. Math. Models Methods Appl. Sci. **29**(4), 717–753 (2019)
62. N. Guglielmi, E. Iacomini, A. Viguerie, Delay differential equations for the spatially-resolved simulation of epidemics with specific application to COVID-19 (2021, preprint). arXiv:2103.01102
63. A.K. Gupta, Models of wealth distributions: a perspective, in *Econophysics and Sociophysics: Trends and Perspectives*, ed. by B.K. Chakrabarti, A. Chatterjee (Wiley, Weinheim, 2006), pp. 161–190
64. H.W. Hethcote, Modeling heterogeneous mixing in infectious disease dynamics, in *Models for Infectious Human Diseases*, ed. by V. Isham, G.F.H. Medley (Cambridge University Press, Cambridge, 1996), pp. 215–238
65. H.W. Hethcote, The mathematics of infectious diseases. SIAM Rev. **42**(4), 599–653 (2000)
66. M. Iannelli, F.A. Milner, A. Pugliese, Analytical and numerical results for the age-structured S-I-S epidemic model with mixed inter-intracohort transmission. SIAM J. Math. Anal. **23**(3), 662–688 (1992)
67. K. Jagodnik, F. Ray, F.M. Giorgi, A. Lachmann, Correcting under-reported COVID-19 case numbers: estimating the true scale of the pandemic (2020, preprint). medRvix:2020.03.14.20036178
68. S. Jin, L. Pareschi, *Uncertainty Quantification for Hyperbolic and Kinetic Equations* (Springer, Berlin, 2017)
69. S. Jin, L. Pareschi, G. Toscani, Uniformly accurate diffusive relaxation schemes for multiscale transport equations. SIAM J. Numer. Anal. **38**(3), 913–936 (2000)
70. S. Jin, D. Xiu, X. Zhu, Asymptotic-preserving methods for hyperbolic and transport equations with random inputs and diffusive scalings. J. Comput. Phys. **289**, 35–52 (2015)
71. S. Jin, H. Lu, L. Pareschi, Efficient stochastic asymptotic-preserving implicit-explicit methods for transport equations with diffusive scalings and random inputs. SIAM J. Sci. Comput. **40**(2), A671–A696 (2018)
72. D. Kahneman, A. Tversky, Prospect theory: an analysis of decision under risk. Econometrica **47**(2), 263–292 (1979)
73. D. Kahneman, A. Tversky, *Choices, Values, and Frames* (Cambridge University Press, Cambridge, 2000)
74. M. Kantner, T. Koprucki, Beyond just "flattening the curve": optimal control of epidemics with purely non-pharmaceutical interventions. J. Math. Ind. **10**(1), 23 (2020)
75. W.O. Kermack, A.G. McKendrick, A contribution to the mathematical theory of epidemics. Proc. R. Soc. Lond. Ser. A **115**(772), 700–721 (1927)
76. A. Korobeinikov, P.K. Maini, Non-linear incidence and stability of infectious disease models. Math. Med. Biol. **22**, 113–128 (2005)
77. D. Lewis, Superspreading drives the COVID pandemic - and could help to tame it. Nature **590**, 544–546 (2021)

78. J.H. Lienhard, P.L. Meyer, A physical basis for the generalized Gamma distribution. Q. Appl. Math. **25**(3), 330–334 (1967)
79. P.L. Lions, G. Toscani, Diffusive limit for finite velocity Boltzmann kinetic models. Rev. Mat. Iberoam. **13**(3), 473–513 (1997)
80. X. Liu, P. Stechlinski, Infectious disease models with time-varying parameters and general nonlinear incidence rate. Appl. Math. Model. **36**(5), 1974–1994 (2012)
81. Y. Liu, A.A. Gayle, A. Wilder-Smith, J. Rocklöv, The reproductive number of COVID-19 is higher compared to SARS coronavirus. J. Travel Med. **27**(2), 1–4 (2020)
82. E. Loli Piccolomini, F. Zama, Monitoring Italian COVID-19 spread by a forced SEIRD model. PLoS ONE **15**(8), e0237417 (2020)
83. S. Loomba, A. de Figueiredo, S.J. Piatek, et al., Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. Nat. Hum. Behav. **5**, 337–348 (2021)
84. N. Loy, A. Tosin, A viral load-based model for epidemic spread on spatial networks. Math. Biosci. Eng. **18**(5), 5635–5663 (2021)
85. Q. Luo, M. Gee, B. Piccoli, D. Work, S. Samaranayake, Managing public transit during a pandemic: the trade-off between safety and mobility (2020, preprint). SSRN:3757210
86. P. Magal, G.F. Webb, X. Wu, Spatial spread of epidemic diseases in geographical settings: seasonal influenza epidemics in Puerto Rico. Discr. Cont. Dyn. Sys. B **25**(6), 2185–2202 (2019)
87. D. Maldarella, L. Pareschi, Price dynamics in financial markets: a kinetic approach. Sci. Culture **76**(9–10), 448–453 (2010)
88. D. Maldarella, L. Pareschi, Kinetic models for socio-economic dynamics of speculative markets. Phys. A Stat. Mech. Appl. **391**(3), 715–730 (2012)
89. K. Mizumoto, K. Kagaya, A. Zarebski, G. Chowell, Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. Euro. Surveill. **25**(10), 2000180 (2020)
90. J.D. Murray, *Mathematical Biology II: Spatial Models and Biomedical Applications*, 3rd edn. (Springer, New York, 2003)
91. G. Naldi, L. Pareschi, G. Toscani, (eds.), *Mathematical Modelling of Collective Behavior in Socio-Economic and Life Sciences*. Modeling and Simulation in Science, Engineering and Technology (Birkhauser, Basel, 2010)
92. B.F. Nielsen, L. Simonsen, K. Sneppen, COVID-19 superspreading suggests mitigation by social network modulation. Phys. Rev. Lett. **126**, 118301 (2021)
93. A.S. Novozhilov, On the spread of epidemics in a closed heterogeneous population. Math. Biosci. **215**, 177–185 (2008)
94. L. Pareschi, An introduction to uncertainty quantification for kinetic equations and related problems, in *Trails in Kinetic Theory: Foundational Aspects and Numerical Methods*, ed. by G. Albi, S. Merino-Aceituno, A. Nota, M. Zanella. SEMA SIMAI Springer Series (Springer, Berlin, 2021)
95. L. Pareschi, G. Toscani, *Interacting Multiagent Systems: Kinetic Equations and Monte Carlo Methods* (Oxford University Press, Oxford, 2013)
96. N. Parolini, et al., SUIHTER: a new mathematical model for COVID-19. Application to the analysis of the second epidemic outbreak in Italy. Proc. R. Soc. A **477**, 20210027 (2021)
97. M. Peirlinck, K. Linka, F. Sahli Costabal, J. Bhattacharya, E. Bendavid, J.P. Ioannidis, E. Kuhl. Visualizing the invisible: the effect of asymptomatic transmission on the outbreak dynamics of COVID-19. Comput. Methods Appl. Mech. Eng. **372**(1), 113410 (2020)
98. M. Peirlinck, et al., Visualizing the invisible: the effect of asymptomatic transmission on the outbreak dynamics of COVID-19. Comput. Methods Appl. Mech. Eng. **372**(1), 113410 (2020)
99. L. Pellis, et al., Eight challenges for network epidemic models. Epidemics **10**, 58–62 (2015)
100. B. Piccoli, M. Garavello, *Traffic Flow on Networks* (American Institute of Mathematical Sciences, Morgan Hill, 2006)

101. G. Poëtte, B. Després, D. Lucor, Uncertainty quantification for systems of conservation laws. J. Comput. Phys. **228**(7), 2443–2467 (2009)
102. K. Prem, A.R. Cook, M. Jit, Projecting social contact matrices in 152 countries using contact surveys and demographic data. PLoS ONE **13**(9), e1005697 (2017)
103. R. Prieto Curiel, H. González Ramírez, Vaccination strategies against COVID-19 and the diffusion of anti-vaccination views. Nat. Sci. Rep. **11**, 6626 (2021)
104. A. Remuzzi, G. Remuzzi, COVID-19 and Italy: what next? Lancet **395**, 1225–1228 (2020)
105. S. Riley, et al., Five challenges for spatial epidemic models. Epidemics **10**(2015), 68–71 (2015)
106. M.G. Roberts, Epidemic models with uncertainty in the reproduction. J. Math. Biol. **66**, 1463–1474 (2013)
107. L. Roques, O. Bonnefon, V. Baudrot, S. Soubeyrand, H. Berestycki, A parsimonious approach for spatial transmission and heterogeneity in the COVID-19 propagation. R. Soc. Open Sci. **7**, 201382 (2020)
108. P.S.A. Salam, W. Bock, A. Klar, S. Tiwari, Disease contagion models coupled to crowd motion and mesh free simulation. Math. Models Methods Appl. Sci. **31**(6), 1277–1295 (2021)
109. E.W. Stacy, A generalization of the Gamma distribution. Ann. Math. Stat. **33**, 1187–1192 (1962)
110. G. Sun, Pattern formation of an epidemic model with diffusion. Nonlinear Dyn. **69**, 1097–1104 (2012)
111. B. Tang, et al., An updated estimation of the risk of transmission of the novel coronavirus (2019-nCov). Infect. Dis. Model. **5**, 248–255 (2020)
112. B. Tang, et al., Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions. J. Clin. Med. **9**(2), 462 (2020)
113. G. Toscani, Statistical description of human addiction phenomena, in *Trails in Kinetic Theory: Foundational Aspects and Numerical Methods*, ed. by G. Albi, S. Merino-Aceituno, A. Nota, M. Zanella. SEMA SIMAI Springer Series (Springer, Berlin, 2021)
114. G. Toscani, Entropy-type inequalities for generalized Gamma densities. Ric. Mat. **70**, 35–50 (2021)
115. P. Van den Driessche, J. Watmough, Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Math. Biosci. **180**, 29–48 (2002)
116. A. Viguerie, et al., Diffusion–reaction compartmental models formulated in a continuum mechanics framework: application to COVID-19, mathematical analysis, and numerical study. Comput. Mech. **66**(5), 1131–1152 (2020)
117. A. Viguerie, et al., Simulating the spread of COVID-19 via a spatially-resolved susceptible–exposed–infected–recovered–deceased (SEIRD) model with heterogeneous diffusion. Appl. Math. Lett. **111**, 106617 (2021)
118. M.A.C. Vollmer, et al., Using mobility to estimate the transmission intensity of COVID-19 in Italy: a subnational analysis with future scenarios. Technical Report of May, Imperial College London (2020)
119. J. Wang, F. Xie, T. Kuniya, Analysis of a reaction-diffusion cholera epidemic model in a spatially heterogeneous environment. Commun. Nonlinear Sci. Numer. Simul. **80**, 104951 (2020)
120. S. Wang, F. Zhong, W. Bao, Y. Li, L. Liu, H. Wang, Y. He, Age-dependent risks of incidence and mortality of COVID- 19 in Hubei Province and other parts of China Hongdou. Front. Med. **7**, 190 (2020)
121. G.F. Webb, A reaction-diffusion model for a deterministic diffusion epidemic. J. Math. Anal. Appl. **84**, 150–161 (1981)
122. D. Xiu, *Numerical Methods for Stochastic Computations - A Spectral Method Approach* (Princeton University Press, New Jersey, 2010)
123. D. Xiu, J.S. Hesthaven, High-order collocation methods for differential equations with random inputs. SIAM J. Sci. Comput. **27**(3), 1118–1139 (2005)

124. M. Zanella, C. Bardelli, M. Azzi, S. Deandrea, P. Perotti, S. Silva, E. Cadum, S. Figini, G. Toscani, Social contacts, epidemic spreading and health system. Mathematical modeling and applications to COVID-19 infection. Math. Biosci. Eng. **18**(4), 3384–3403 (2021)
125. M. Azzi, C. Bardelli, S. Deandrea, G. Dimarco, S. Figini, P. Perotti, G. Toscani, M. Zanella, A data-driven epidemic model with social structure for understanding the COVID-19 infection on a heavily affected Italian Province. Math. Models Methods Appl. Sci. **31**, 2533–2570 (2021)
126. S. Zhang, M. Diao, W. Yu, L. Pei, Z. Lin, D. Chen, Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: a data-driven analysis. Int. J. Infect. Dis. **93**, 201–204 (2020)
127. D. Zhang, M. Hu, Q. Ji, Financial markets under the global pandemic of COVID-19. Finance Res. Lett. **36**, 101528 (2020)

The COVID-19 Pandemic Evolution in Hawai‘i and New Jersey: A Lesson on Infection Transmissibility and the Role of Human Behavior

Sarah Allred, Monique Chyba, James M. Hyman, Yuriy Mileyko, and Benedetto Piccoli

1 Introduction

The COVID-19 pandemic has caused an unprecedented disruption to the world, infecting millions and shutting down large parts of the world economy. We illustrate the value of mathematical epidemiological models in understanding the current pandemic and discuss how a team-science approach to modeling may effectively prevent and mitigate the effects of future pandemics. After outlining the challenges confronting the early modeling of the COVID-19 pandemic, we provide a broad overview of agent-based models (ABMs) and compartmental models (CMs). We then use the models to simulate the epidemic in Hawai‘i and New Jersey and discuss what could be done to prepare for the next pandemic.

Mathematical models have been central to the public and policy debate throughout the COVID-19 pandemic. When combined with empirical data, modeling provides a valuable tool to:

S. Allred
Department of Psychology, Rutgers University–Camden, Camden, NJ, USA
e-mail: srallred@camden.rutgers.edu

M. Chyba · Y. Mileyko
Department of Mathematics, University of Hawai‘i at Manoa, Honolulu, HI, USA
e-mail: chyba@hawaii.edu; yury@math.hawaii.edu

J. M. Hyman
Department of Mathematics, Tulane University, New Orleans, LA, USA
e-mail: mhyman@tulane.edu

B. Piccoli (✉)
Department of Mathematical Sciences, Rutgers University–Camden, Camden, NJ, USA
e-mail: piccoli@camden.rutgers.edu

N. Bellomo, M. A. J. Chaplain (eds.), *Predicting Pandemics in a Globally Connected World, Volume 1*, Modeling and Simulation in Science, Engineering and Technology, https://doi.org/10.1007/978-3-030-96562-4_4

- characterize the extent of an epidemic,
- predict the scale of the healthcare problem, and
- explore the impact of different intervention strategies.

Models rely on the parameters coded into the model and the data used to constrain those parameters. These parameters are estimated based on the best available data. The uncertainty in the data limits our ability to quantify the predictive model forecasts. In a novel pandemic, the quality of the data available is constantly changing, and the model predictive accuracy improves as more empirical data becomes available. Even the best-developed models are unlikely to be accurate prediction tools over a long time. Unfortunately, the mismatch between the predictions of many initial models and the ultimate outcomes of the pandemic brought heavy criticism about their purpose and credibility.

In the COVID-19 pandemic, the relative importance of the model parameters to the transmission has changed over time to account for changes in social distancing mask-wearing and vaccinations. Also, early in the pandemic, it was unclear how efficiently the virus could be transmitted through air, fluids, or surfaces. It took time to accumulate the data needed to characterize the distance and time scales over which such transmission could occur. Similarly, the role of asymptomatic cases is still not well quantified more than a year into the pandemic. With this confusion about basic empirical transmission facts, there has been no clear methodology to accurately estimate some of the models' most important parameters.

The early transmission models relied on a limited understanding of the importance and role of airborne transmission of the virus. Also, most scientists developing the models only had access to publicly available aggregated data. This available data

- lacked the needed granularity for the age, symptoms, or hospitalization of the samples;
- often combined the results of the antigen and antibiotic tests; and
- lacked information about the bias in the testing so the results can be used to estimate the incidence. For example, the available data rarely included the symptoms, malaise, and age of each tested individual.

Mathematical models based on unreliable data will provide inaccurate forecasts and run the risk of being inappropriately used. The rapid spread of the infection created an emergency where we lacked reliable estimates for the information needed to make quick decisions in an acute setting. Often the most reliable information came from asking a group of experts and averaging their opinions.

On top of these challenges, the evolution of a pandemic is people-driven, and human behavior is complex, unpredictable, and changes with time, possibly very rapidly. Social distancing measures and masking measures at the macro-scale of humans are not well understood, and reliable estimates do not exist. Human behavior in epidemiological models must include considerations on the type of interactions between individuals. The models must account for how the infection spreads more rapidly in social gatherings and household environments than in

general settings. They can quantify the links between restaurants, social groups, and churches to localized outbreaks. They can estimate how an epidemic will disproportionately impact multi-generational families than single-family units.

Although most of the available data had severe limitations, there were constantly updated databases tracking the spread of COVID-19 in North America and the European Union. The lack of almost any information from third-world countries created huge challenges to understand the global scale of the COVID-19 pandemic.

One of the most informative model parameters is the transmissibility of the infection, as measured by the *basic reproductive number*, $\mathcal{R}_0$. The basic reproductive number is the average number of people a newly infected person will infect in a population where everyone else is susceptible. This depends on the behavior of the population as measured by the number of contacts a typical person has each day, the infectiousness of the virus (which depends on social distancing and mask-wearing), and the length of time a person is infectious. Although it might be reasonable to assume $\mathcal{R}_0$ is fixed during a typical flu season, the dramatic behavior changes in response to the COVID-19 pandemic require the models to account for these changes and vary $\mathcal{R}_0$.

The early COVID-19 models assumed the parameters were static, meaning that most parameter values, and hence $\mathcal{R}_0$, were fixed. Given the sparse data available and the limited understanding of virus transmission, these models were useful to explore the virus's possible outcomes but had low quantitative accuracy. They were helpful in illustrating the "Flatten the curve" concept among the public and policy leaders. Unfortunately, as models developed both to become dynamic and incorporate additional knowledge about COVID-19, the lack of quantitative accuracy of initial models resulted in a backlash for the modeling.

As the virus spreads and people become immune to infection, either by recovering from an infection or being vaccinated, the fraction of the population susceptible to infection, P_s, will decrease. A typical infected person will only infect the *effective reproductive number* $\mathcal{R}_e = \mathcal{R}_0 P_s$ of people. That is, R_e represents the number of people that a newly infected person will infect during an epidemic. If $R_e < 1$, then the population has reached *herd immunity* and the epidemic will die out.

The effective reproductive number, and therefore herd immunity, depend on both P_s (vaccination and recovered infected) and R_0, which includes social distancing, mask-wearing, and other behavior that affects the virus's transmissibility. A population with herd immunity while social distancing can lose it if the population relaxes social distancing and $\mathcal{R}_0$ increases. Similarly, as more and more people are vaccinated, P_s will decrease until $\mathcal{R}_e = \mathcal{R}_0 P_s < 1$ and herd immunity is achieved.

The COVID-19 pandemic will continue until there is sufficient worldwide vaccination to achieve herd immunity. There will be infection waves due to geographically sparse outbursts, the appearance of multiple variants, and diverse political and social responses to the pandemic. Despite this somehow unfortunate reality, models can help inform decision-makers to manage the current pandemic and prepare for the next one. Globalization will continue to connect the modern world, with large masses traveling throughout the globe.

Recent COVID-19 containment measures have included travel bands, testing, and quarantining the sick and exposed populations. Measures used sporadically in the current epidemic have included:

- Taking temperatures in restaurants, workplaces, and travelers at start and end of travel;
- Require testing before take-off and at landing for long-distance air travels;
- Quarantining after arriving in a new location, such as another country or state; and
- Contact tracing for people in contact with known infected individuals.

All of these mitigation efforts reduced $\mathcal{R}_e$, but, unfortunately, many of these measures were relaxed in the US, just as the COVID-19 Delta variant surged across the continent.

Researchers have recently focused on including behavior changes in the models to improve their predictive capabilities and identify which activities could most effectively mitigate the pandemic. Although some of the current models include these behavior changes, the lack of reliable data for human behavior limits their predictability. Also, the behavior of a single infected individual can affect the spread of an infection, especially in the early stages of an epidemic. Once the infection becomes widespread, data for the aggregated community-level behavior might be sufficient for the models to predict the spread reliably.

Predicting the transmission pathways that an infection can take depends on the decisions people make for social distancing, masking, self-reporting symptoms, testing, quarantining and isolation, and vaccination. A new approach called infodemic modeling simulates the spread of good and bad information about the epidemic and predicts how the public will respond. These models include the public's fatigue to restrictive measures, confusion around changing policies, the role of misinformation on social media.

Transmission models for the COVID-19 pandemic must be complex to capture the complexity of the changes in people's behavior and vaccination status. Effective models require collaborating teams of mathematicians, social scientists, and academic and government public health experts to build the simulations on a solid mathematical and epidemiological foundation.

We illustrate this complexity by comparing the spread of the virus in Hawai‘i, with natural geography barriers between the major cities, and the spread in New Jersey, where population centers are tightly connected within the state and the neighboring major cities, New York and Philadelphia. In addition, New Jersey can be divided into three distinct regions with different sociodemographic characteristics. We use our models to help understand why the epidemic evolved very differently in these two regions.

After describing the advantages and limitations of agent-based and compartmental transmission models, we fit the models to simulation the epidemics in Hawai‘i (Sect. 3) and New Jersey (Sect. 4). We relate the magnitude of the multiple infection waves to the geographical isolation of the Hawaiian islands and the sociodemographic characteristics of New Jersey. We observed that our CM for New

Jersey (Sect. 5) was effective in predicting the early stage in the epidemic but less accurate as the pandemic evolved.

2 Mathematical Models

Agent-based models (ABMs) mimic the behavior of a group of individuals or agents and their interactions (contacts) where the infection could be transmitted from an infected person to a susceptible person [37]. Often the model can be represented by a network or graph where the nodes represent the individuals. The edges between the nodes represent contacts where an infection could be transmitted from an infected to a susceptible person. The infection transmission is modeled as a stochastic event, and each epidemic simulation tracks a random branching event as the infection spreads through the network.

An ABM can be compared to modeling fluid dynamics by following the paths of individual atoms. One could imagine modeling the Brownian motion of multiple particles that all follow the same physical laws. As the number of particles increases, eventually, the average behavior of the system can be predicted by a mean-field diffusion equation. In the same spirit, as the number of infected individuals in an ABM becomes widespread, then the average spread of infection can be modeled by a mean-field CM.

Compartmental epidemiological models are based on the idea of splitting the population into different compartments corresponding to their behavior or infection status. The underlying assumption is that all of the individuals within a compartment are indistinguishable.

2.1 Agent-Based Models

ABMs have been used in social, economic, and biological sciences for over 40 years [46, 51, 52]. The popularity of ABMs exploded in the 1990s when the computational power significantly increased and became widely available, and they have been used to model the spread of influenza, measles, Ebola, and other infectious diseases [3, 9, 10, 27, 32, 41, 47].

The connectivity properties of an ABM network, which is described in more detail below, reflect the contact network in the particular population being modeled and impact the spread of the epidemic. These properties are quantified by the degree distribution, joint-degree distribution, and clustering coefficient. While an individual behavior can often be described using simple rules, the collective behavior of the network of individuals can exhibit complicated interactions and transmission pathways. Each epidemic can follow a wide range of paths depending on the structure of the network. The uncertainty generated by the stochastic transmission

network creates challenges when interpreting the results to non-statisticians and policymakers.

In an ABM, the state of each agent can be described as an element of some space S. If A is the finite set of agents, then the state of the ABM system can be defined by a map of each element in A to an element in S. The collection of all such maps denoted S^A, constitute the total state space of an ABM. When modeling a pandemic, one typically chooses A to be the set of people within a closed community of interest, and S is usually a set of tuples consisting of continuous and binary values. Formally, such a set of tuples can be written as a product $\mathbb{R}^m \times \mathbb{Z}_2^k$. Here $\mathbb{R}^m$ captures continuous variables related to the disease, such as age and susceptibility to infection, and $\mathbb{Z}_2^k = \{0, 1\}^k$ captures binary variables, such as the presence of infection and comorbidities. Of course, one can add other categorical variables if needed.

The evolution of the system state through time is described by an association of an element of our total state space to each time point under consideration. Again, such an association can be regarded as a map from the collection of possible times to the total state space of the system, which we can formally write as $g : T \to A^S$. Here, T is our time set, which, assuming that the state changes at discrete time points, can be regarded simply as the set of natural numbers, $T = \mathbb{N}$. Instead of looking at the total state of the system, we can, equivalently, focus on the evolution of the state of each individual agent. This means that we associate an agent's state, i.e., an element of S, to each time-agent pair.

Formally, this yields a map $f : T \times A \to S$, which is related to the evolution of the total state via the equality $f(t, a) = g(t)(a)$. The function f can be defined deterministically, or it can be a realization of a stochastic process. Importantly, one does not focus on the whole f and instead defines how the state of an individual agent is updated using the states of the other agents. That is, one defines how the value $f(t + 1, a)$ is obtained when $f(t, a)$ for all $a \in A$ is known. Of course, a change in the state of an agent is unlikely to depend on all of the other agents. Typically, each agent has an associated subset of agents that may affect its state due to "interaction". We shall refer to such a subset as a *contact set* of an agent. In the case of pandemic modeling, a contact set consists of people who actually interact with a given individual. This example also suggests that a contact set of an agent $a \in A$, which we shall denote $N(a)$, may have an additional structure to better reflect interactions within different contexts. For example, interactions at work may be different from those at home. Mathematically, this may be represented as a disjoint union, $N(a) = \sqcup_{i=1}^n N_i(a)$. Also, a contact set may be time dependent, thus yielding $N(t, a) = \sqcup_{i=1}^n N_i(t, a)$, $t \in T$, which simply says that the contact set may change with time.

Considering that interactions between agents are typically symmetric, it is convenient to represent all contact sets $N(a)$, $a \in A$, using an undirected graph, so that each separate $N(a)$ is just a collection of adjacent vertices. More specifically,

we let A be the vertex set of our graph, and let $E \subset \{\{a, b\}|a, b \in A\}$ be the set of edges. We shall refer to such a graph as a *contact network*. We can then define a contact set of an agent $a \in A$ as a set of adjacent vertices, i.e., $N(a) = \{b \in A|\exists\{a, b\} \in E\}$. Additional structure and time dependency of contact sets are obtained by defining multiple, possibly time-dependent interaction networks with the same vertex set A and different edge sets $E_i(t), i = 1, \ldots, n$.

In the deterministic case, the state of an agent $a \in A$ at time step $t + 1$ is defined as a function of time t and the states $f(t, b)$ where b ranges over $N_i(t)$, $i = 1, \ldots, n$. In the stochastic case, one computes the conditional probability distribution for $f(t + 1, a)$, conditioned on the above $f(t, a)$ (and possibly t), and then samples from it. When modeling a pandemic, this often simplifies computing the probability of infection given an uninfected individual or the probability of developing symptoms, given an infected but asymptomatic individual, etc. Then a (pseudo)random number is generated to determine the actual state transition (e.g., an individual gets infected, develops symptoms, etc.). While the transition between states can be mathematically quite complicated and non-Markovian, the implementation is often fairly straightforward.

An example of a contact network for an ABM of a pandemic is shown in Fig. 1. The contact network determines how the infection will spread through the community, and their construction is a challenging problem. Once the network is

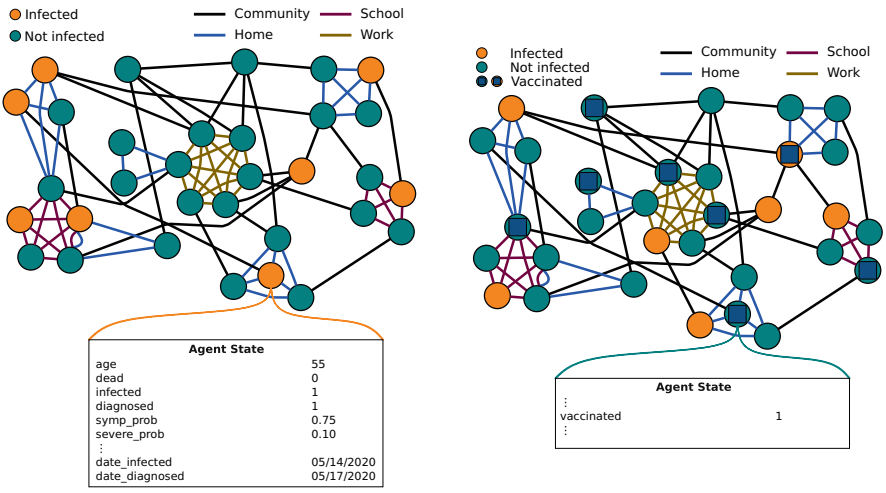


Fig. 1 Sample contact network representing individuals in the population as nodes and the interactions for possible viral transmission among them as edges. The different colors refer to four different types of contacts or individuals in the population (LHS). The vaccinated individuals will have a reduced transmission which is reflected in their state (RHS)

defined, the infection is transmitted through the edges (contacts) between the nodes (people), and the infection status can be straightforwardly updated. For example, Fig. 1 shows that adding vaccinated individuals is conceptually quite simple. Of course, the process governing state transitions also needs to be updated.

2.1.1 COVID-19 Agent-Based Simulator (Covasim)

Covasim is a stochastic ABM developed by the Institute for Disease Modeling to explore the coronavirus dynamics and interventions [35]. The Covasim default contact networks have our "social layers" (i.e., edge sets) analogous to the ones shown in Fig. 1. Covasim associates each agent (node) a state with 39 variables, including information such as demographic, individual susceptibility, and variables representing intrahost viral dynamics (along with viral-load-based transmissibility). Each stochastic simulation generates a different epidemic. The expected values for quantities of interest (e.g., new daily infections, the number of hospitalizations, etc.) are reported for an ensemble of (typically 15–20) simulations.

We extended Covasim to include additional interventions, such as vaccinations, for our simulations of the Hawaiian COVID-19 epidemic in Sect. 3.1.1.

2.2 Compartmental SEIR Models and Variants

Compartmental epidemic models started with the pioneering Kermack-McKendrick Susceptible-Infected-Recovered (SIR) model [34]. The susceptible compartment (S) included individuals who could be infected if they are in contact with an infected person (I). The infected population eventually recovered (R) and were no longer susceptible to being infected.

The original SIR model has been extended to include additional states, such as including an infected, but not infectious, exposed state (E) to create an SEIR model [31]. Further extensions have proved effective in modeling traditional infectious diseases (pertussis [20, 48], measles [5], Ebola [38, 40], etc.), new diseases (HIV [21, 45], Lyme, etc.), as well as reemerging diseases (tuberculosis [6, 54], cholera [19], etc.). They have been used extensively also for the COVID-19 pandemic [7, 23, 25, 26, 63].

For example, a simple CM was used for predicting hospital beds that would be needed in New Jersey in the early stages of the COVID-19 epidemic [1]:

$$
\begin{aligned}
\frac{dS}{dt} &= -(I_A + I_S)\mathcal{R}_0 \left(\frac{S}{N}\right)\frac{1}{D_I} \\
\frac{dE}{dt} &= (I_A + I_S)\mathcal{R}_0 \left(\frac{S}{N}\right)\frac{1}{D_I} - \frac{E}{D_E} \\
\frac{dI_A}{dt} &= \alpha\frac{E}{D_E} - \frac{I_A}{D_I} \\
\frac{dI_S}{dt} &= (1-\alpha)\frac{E}{D_E} - \frac{I_S}{D_I} \\
\frac{dH}{dt} &= \sigma\frac{I_S}{D_I} - \frac{H}{D_H} \\
\frac{dR}{dt} &= (1-\sigma)\frac{I_S}{D_I} + \frac{H}{D_H},
\end{aligned}
\tag{1}
$$

where S are susceptible, E exposed, I_A infected asymptomatic, I_S infected symptomatic, R recovered, and $N = S + E + I_A + I_S + R$ is the total population.

The basic reproduction number, $\mathcal{R}_0$, is the average number of individuals infected by infected individual *if* everyone is susceptible. During an epidemic, only the fraction $P_s = S/N$ of the population is susceptible, and so a typical infected person will only infect the effective reproductive number $\mathcal{R}_e = \mathcal{R}_0(S/N)$ people. If an individual is infected for an average of D_I days, then, on average, an infected individual will infect $\mathcal{R}_e/D_I$ people per day. Hence, the total infected population, $(I_A + I_S)$, will infect $(I_A + I_S)\mathcal{R}_e/D_I$ people per day. All the newly infected people are removed from the S compartment in the first equation and enter the exposed, E, compartment.

The exposed population becomes infectious after an average of D_E days when a fraction α enters the asymptomatic infectious state I_A and $(1 - \alpha)$ enter the symptomatic infectious state. The model assumes that the infectious people either recover or are hospitalized after an average of D_I days. A before fraction σ of I_S are hospitalized before recovering after D_H days.

This simplified model assumes a closed, randomly mixing population where the small fraction of people who die from the infection do not significantly impact the overall spread of the infection. Even with these simplifications, this model helped estimate hospital bed needs in the initial phases of the pandemic. However, the model assumed that everyone had the same behavior and risk of infection, and the coarse-grain model could not be used to forecast the spatial heterogeneities in the epidemic [36].

There are more complicated CMs that include time-dependent parameters, effect of travels and even zoonotic infections [11, 50]. These more complicated models have more compartments for different progressions of the disease [26] and account for age-structured populations and spatial variability [8, 16, 17, 33, 61, 62].

The simulations in Sect. 3 for the spread of the virus in Hawai‘i generalize the model by Lloyd-Smith et al. [39] and subdivide the susceptible population into the general community (C) and healthcare workers (H). In addition, Exposed and Infected (in each population group) are split into multiple stages per day to reflect the disease's progression better. Also, individuals in isolation (including hospitalization) are categorized in separate compartments. The incubation period for those who do develop symptoms is between 2 to 14 days after exposure. The model assumes that some of the people who contract the virus are asymptomatic. Those who do not develop symptoms after 14 days are moved from the Exposed compartment to the Removed compartment.

The quarantine sub-compartment has 14 stages accounting for the effect of contact tracing and the reduced transmission rate for quarantined individuals. The infected individuals go through 5 stages, of which the first two represent the first two days of being symptomatic, whereas the last three represent the phase where the immune system is fighting the disease. Since the last three stages can go on for more than one day each, there is variability in the number of days any given person can spend at each stage. Our model assumes that the symptomatic phase of the illness lasts at least five days.

Travelers play an essential role in spreading the infection in the archipelago and Islands. We consider two broad categories of travelers—tourists (forming a new group) and returning residents (coming into the community group). Finally, the Vaccinated population has the same sub-structure as the Exposed and Infected compartments used for non-vaccinated groups, with the parameters modified to account for the vaccination status. This CM is described in detail in [37].

2.3 Comparison of Agent-Based and Compartmental Models

For the past two decades, there has been a robust discussion on the effectiveness of ABM and CM epidemic models [42, 43, 58]. ABMs and CMs are structurally different. Agent-based models treat the population at the individual level, where every individual can have a different behavior pattern or disease progression. Compartmental models aggregate the population into compartments, where the people within the same compartment are indistinguishable. Stochastic ABMs can capture the early and late stages of an epidemic when the course of an epidemic could be sensitive to individual behavior and transmission events. However, ABMs are computationally expensive and challenging to analyze when an ensemble of simulations predicts a broad range of possible epidemics. Ordinary differential equations approximate the mean-field behavior of a population and best once the infection has become well-established. The analysis of CMs is straightforward, and simulations are inexpensive.

Kunwar et al. [37] summarize the advantages and limitations of both models. When dealing with a CM, an inherent difficulty is the conceptual complexity associated with accounting for newly discovered pandemic features, such as vaccines

and new variants, which can be highly non-trivial to implement. The aggregate nature of the interactions poses a challenge to reflect the actual intricacy of those interactions. Therefore, the conceptual complexity of the model may become a hurdle. A CM is also quite limiting when it comes to including demographic, ethnic, and other essential information, as it would again require one to introduce numerous compartments. On the other side, due to the structure of ABMs and everyone being represented by an agent, incorporating new attributes for individuals, such as age, ethnicity, or vaccination status, is conceptually relatively simple. It amounts to augment the dimension of the state of each agent while the network is untouched. For instance, a new variant will be simply represented as a new variable characterizing the variant to the agent state space.

Stochastic network ABMs are computationally expensive when there are many agents (nodes) and widespread infection. Fortunately, it is often straightforward to parallelize ABMs, and current software can often handle populations with tens of millions of people. Usually, only a tiny fraction of the population is infected at any given time in an epidemic. Agent-based transmission models only need to follow this small, infected population to follow the paths of infection. For example, if there are $|A_I|$ infected nodes (people) and there are, on average, $|E_I|$ edges (contacts) per node, then there are at most $|A_I|\,|E_I|$ people who could be infected on the next time step. Note that this is independent of the size of the network.

Although the transmission dynamics requires $O(|A_I|\,|E_I|)$ operations, when the behavior or the vaccine status of population is changing, then each time step may require $O(|A|)$ operations to update the status of all $|A|$ agents and $O(|E|)$ operations if the contact structure changes in the number of edges, $|E|$, due to changes in social distancing.

The primary challenge for the ABMs is that they might be computationally expensive if the number of agents (i.e., the population) is large and the computation is not appropriately parallelized. Since every time step requires an iteration over each interaction edge and an update on each agent's state, the size of the network impacts primarily the total time complexity, which is $O(|E| + |A|)$ of each time step where $|E|$ is the number of interaction edges and $|A|$ denotes the number of agents. In [37] the authors show that the increase in computational time with the number of days is roughly linear and that the scaling of the computational time with respect to the number of agents is also close to being linear. The role of parallelism in reducing the computational time of agent-based epidemiological models should be explored to speed up calculations crucial for large population sizes. On the other side, the computational complexity is typical for CMs $O(NC)$, where NC is the number of compartments, which is significantly less computationally intensive for the same problem.

Agent-based and compartment models complement each other and are often best used together. For example, classical gradient-based optimization algorithms can easily fit the parameters for a CM to data and determine which parameters are identifiable. It is, however, far more complicated to estimate parameters or determine their identifiability using an ABM. As explained in [37], often you need to resort to a stochastic global optimization technique such as the Metropolis-Hastings

algorithm or the genetic algorithm [12, 29]. Such a procedure is computationally expensive and does not always produce a good fit [29]. We can combine the approaches by using the CM to obtain reasonable estimates for the baseline parameters. These values can then be used as an initial guess (or a prior) to calculate appropriate values of the corresponding parameter in an ABM.

3 Archipelagos and Islands

Little research has focused on modeling pandemics in smaller, contained populations. It is, however, imperative to study and understand the spread of infectious diseases in such environments since populations in them are highly vulnerable.

We will use the Hawaiian Islands to illustrate the challenges and benefits of a contained environment. Documentation on the impact of infectious diseases in Hawai‘i dates back to the eighteenth century. It exposes how, due to lack of immunity to foreign diseases, the native Hawaiian population has been decimated in the past by invasive viruses such as measles, whooping cough, dysentery, and influenza. The first quarantine facility was established by King Kamehameha V in Molokai for individuals affected by leprosy. In addition to many Native Hawaiian lives lost, schools and the economy were disrupted due to the high mortality [4, 30, 53]. Today, Hawai‘i remains vulnerable to infectious diseases because it is geographically isolated and does not have hospitals and other medical facilities capable of treating large numbers of infected non-residents; nor does it have the ability to shift excess COVID-19 cases to neighboring hospitals or care centers other than perhaps military ships.

The State of Hawai‘i found itself in a unique position in the COVID-19 pandemic due to its isolated geographic location, mostly linear population distribution along the coast, and a heavy dependence on tourism and hospitality sectors of the economy [14]. At the time of writing this paper, early September 2021, Hawai‘i is facing its biggest crisis since the pandemic began. There is a clear dichotomy between the conditions before and after the introduction of the Delta variant. What remains true, however, is that despite their proximity and similar environment, different Hawaiian counties exhibit quantifiably different dynamics of the spread of the disease [15] and therefore require localized, targeted mitigation measures, possibly at different times.

3.1 March 2020–June 2021

Hawai‘i ’s first case of COVID-19 was detected on March 6, 2020, in a man that traveled back to the Islands from a cruise. A sharp exponential growth followed the traveler’s return, and a stay-at-home order was declared on March 25, 2020. A CM was developed during the summer of 2020 to evaluate the impact of various

mitigation measures and, once calibrated to the observed case numbers, forecast what-if scenarios. This model, described in [13], quantifies the differences and similarities in the spread of COVID-19 among the different Hawaiian counties in 2020. The authors compiled reported COVID-19 daily cases from March 6, 2020, to January 15, 2021, grouped by the four primary counties: Honolulu, Hawai‘i, Maui, and Kaua‘i. They examine available data on daily cases, positivity rates, mobility and employ the CM mentioned above fitted to the daily cases. We first summarize the results presented in that paper.

The transmission rate in their CM model is optimized to reflect non-pharmaceutical interventions. These included the two stay-at-home orders early in the pandemic, as well as the safe travel program implemented on October 15, 2020. Fall 2020 also saw an introduction of the Tier system in Hawai‘i based on the count of daily new infections and testing positivity rate. The State moved from Tier 1 to Tier 2 on October 21, 2020, from Tier 2 to Tier 3 on February 25, 2021, and then from Tier 3 to Tier 4 on June 11, 2021. Unique to a geographically isolated location is the ability to control travel. The State of Hawai‘i has heavily regulated travel; see [13] for more details. The travel flux in the model is based on data compiled from the COVID-19 State of Hawai‘i Travel portal [57]. The State of Hawai‘i started administering vaccines on December 15, 2020. The model vaccination rate was estimated from the State of Hawai‘i—Department of Health, Disease Outbreak Control Division Dashboard.

3.1.1 CM Model Fit from March 06, 2020 to January 15, 2021

Figure 2 displays the model fit for the Honolulu county. The dots represent the new daily cases reported by the State of Hawai‘i Department of Health, and the red curve shows the model fit. The vertical lines correspond to mitigation measures that had an impact on the curve and thus governed the values of the parameter representing the baseline transmission rate, β. The model was fit to the data by varying the transmissibility, β (shown with the red bar graphs). The rapid decrease in β at the beginning of April demonstrates the significant impact of the first stay-at-home order on March 25 while numbers were still under 50 new daily cases.

On August 27, 2020, the second stay-at-home order had a lesser impact on the transmission rate due to the ongoing community spread within large households. On October 15, 2020, the safe travel program was initiated with mandatory pre-travel testing for all passengers or a 2-week quarantine. It was followed by a surge in travel and initiated an increase in tourism which has effectively become nonexistent from March 20 to October 15, 2020. As seen in Fig. 2, in Honolulu County, the new daily cases stabilized between 70 and 100 during Fall 2020, with a slight increase due to the Holidays.

An analysis of the other counties showed significant differences among them in the evolution of the pandemic. On Fig. 3, obtained from [13], we display the normalized model for Honolulu, Hawai‘i, and Maui counties as well as the daily raw numbers for Kaua‘i (there is no model fit for Kaua‘i due to the numbers being

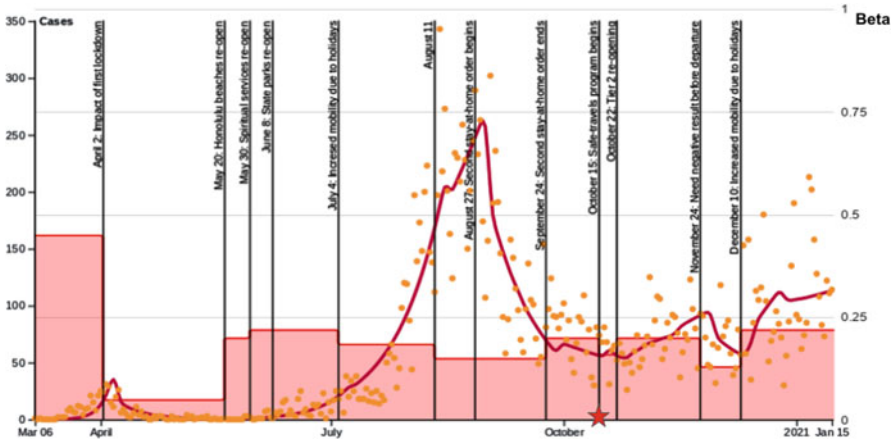


Fig. 2 Honolulu County model fit from March 06, 2020 to January 15, 2021. The star shows the beginning of safe travel program. It can be observed that the compartmental model accurately tracks the data

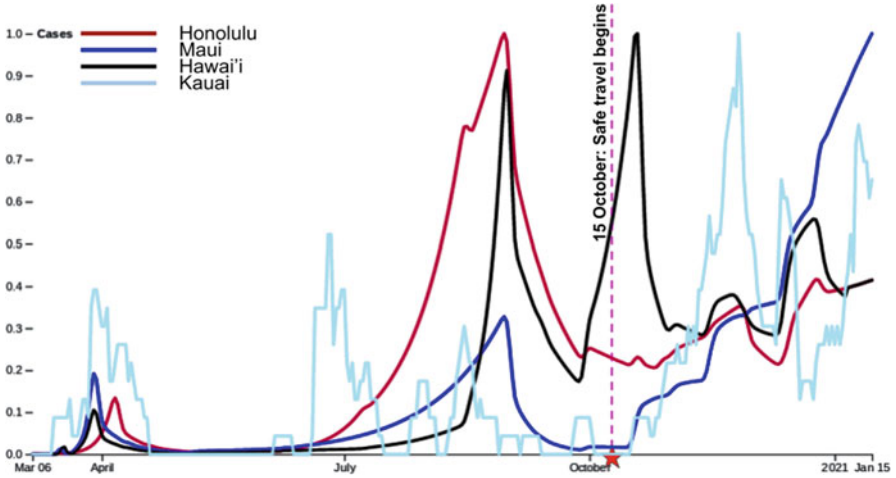


Fig. 3 Honolulu, Maui, and Kaua‘i counties with the normalized model fit, Kaua‘i with normalized daily cases. It is clearly observed that counties started to differ in response to the spread of COVID-19 after the safe travel program opened

too small). Except for Kaua‘i, the other three counties exhibited qualitatively similar behavior until the Safe Travel program began. Hawai‘i and Maui county had a few clusters that generated the other peaks. The main observation is the sharp increase in new daily cases for Maui starting end of October, which is not visible on the other Islands.

We analyze similarity between the counties of Hawai‘i by computing the classical L_2 distance between normalized model fits. Considering that we have a

Table 1 L_2 norm between normalized fit of Hawaiian counties (Fig. 3) to measure similarity. The counties have different tourist-to-resident ratios, and the travel restrictions impact the counties differently. Before travel was reinstated, Honolulu and Maui counties were the most similar; this situation changed afterward

	Mar 06–October 15	Mar 06–Jan 15
Honolulu–Maui	2.05	4.50
Honolulu–Hawai‘i	3.31	3.81
Hawai‘i–Maui	3.53	4.23

single sample for each of the time series representing the new daily cases, we assess the significance of the computed L_2 distances by looking at how the corresponding numbers change when the model is fitted to random perturbations of the original time series data. These perturbations are obtained by adding Gaussian noise with mean zero. Assuming that a large number of new daily cases implies a larger possible error, the standard deviation for the noise is time dependent and equal to one-tenth of the recorded daily cases. In the rare cases when adding the noise yielded a negative number of new daily cases, this number is set to zero. In Table 1 we compare the L_2-norm of the COVID-19 incidence in the different counties.

The work presented in [13] also compares the different counties to other Islands and Archipelagos. Using the merged trees corresponding to the new daily cases curves shows that Honolulu county is qualitatively most similar to Iceland, Maui county to Japan, and Hawai‘i county to Puerto Rico (to a smaller degree). The conclusion is that islands with control over the influx and outflow of people may surprisingly experience similar overall pandemic experiences, even if they are geographically distant and culturally different. This suggests an underlying complex transmission network for the spread of the disease within an island chain, since the similarity in weather and proximity do not always correlate with the spread of the infection.

In December 2020, the state of Hawai‘i began vaccinating people. The CM results, presented below, come from [37] assume that the vaccine requires two doses. New compartments were introduced into the model to account for vaccination and differentiate between people who have received one or two doses of the vaccine. The model also tracks exposure to the virus and infections occurring in vaccinated individuals between vaccine doses and post-vaccination. The vaccinated compartments have the same sub-structure as the Exposed and Infected compartments used for non-vaccinated groups, but parameters might vary. The interaction between the vaccinated and non-vaccinated individuals is accounted for by the hazard rate for a susceptible person's likelihood of getting infected. Figure 4 shows the dynamics of the model that includes travelers and vaccines for Honolulu county. The simulations assume an 80% reduction in transmissibility for the vaccinated population and a 95% reduction in the probability that a vaccinated individual will develop symptoms and severe disease. The parameters values for these simulations are in [37].

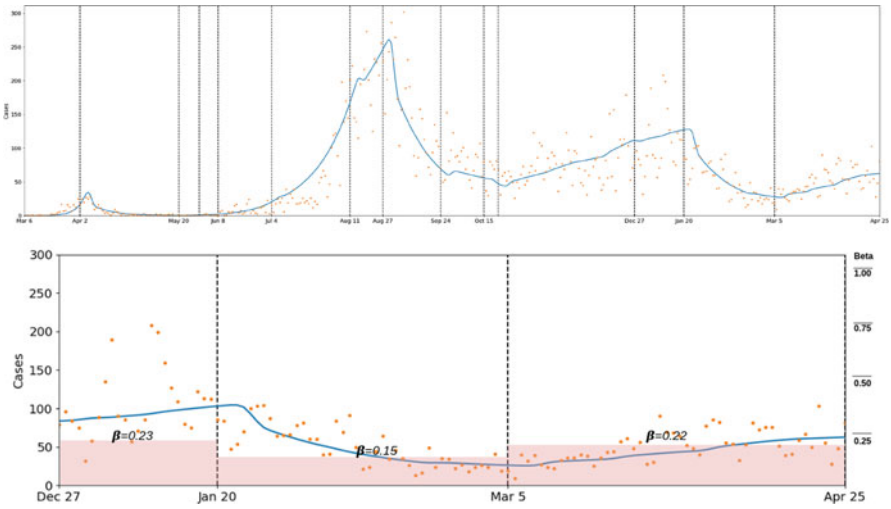


Fig. 4 Top: Honolulu County fit from March 6, 2020, including travelers starting October 15, 2020, and vaccination starting December 27, 2020. The simulations through April 25, 2021. Bottom: Zoom in for the period where both vaccination and travelers are included with the corresponding transmission rates

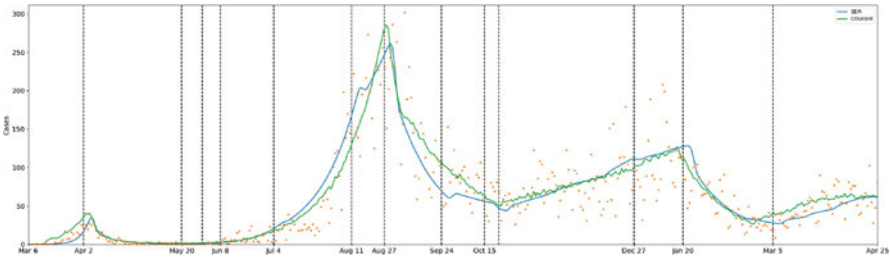


Fig. 5 Compartmental (blue) and Covasim (green) fit from March 6, 2020 to April 25, 2021 including travelers and vaccines [37]

3.1.2 Comparing CM and ABM Models

In Fig. 5, we represent the optimized fit obtained from the Compartmental SEIR (blue) and the agent-based Covasim (green) models for Honolulu County. To run Covasim, we used the data from the Hawai'i Population Model developed by the Hawai'i Data Collaborative [28]. It can be observed that the two models agree qualitatively, including the spike in August.

Figure 6 shows the impact of the vaccines for both models with the benchmark fitted parameters. The simulation is for 1 million people with 100 infections on day 1 (but no symptoms) and constant transmissibility, β. Scenario 1, shown in Fig. 6 top,

assumes no vaccines and compares the model predictions over 425 days. Scenario 2 assumes a vaccination rate of 2500 individuals per day and a 95% reduction in the probability of developing symptoms for vaccinated individuals and an 80% reduction in transmissibility. The simulations also assume that the vaccine only requires a single dose. The results are shown in Fig. 6 bottom.

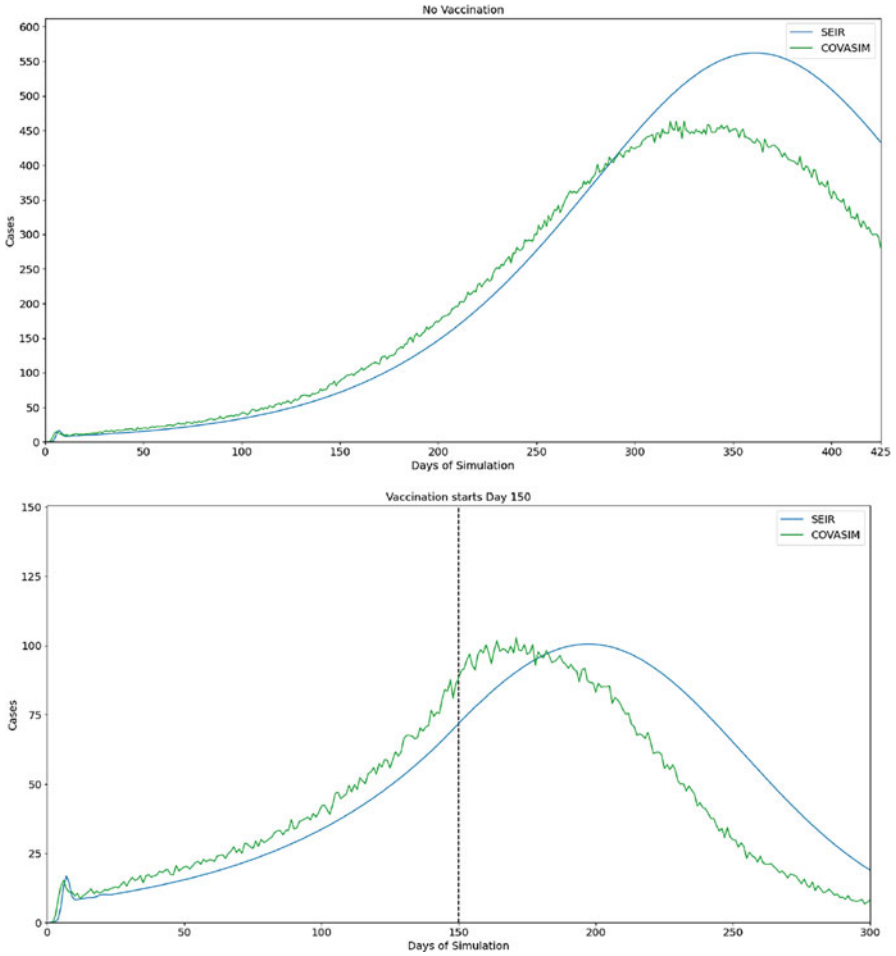


Fig. 6 SEIR (blue) and Covasim (green) benchmark scenarios forecasting spread and vaccination. The green Covasim curve is the mean of 20 simulations [37]. Observe that the overall behavior of the two models is similar even without fitting the data

3.2 July 2021–September 2021

The first case of Delta variant in the State of Hawai‘i was detected in a fully vaccinated Oahu resident who had traveled to Nevada in early May. As of June 13, 2021, it was present in all four counties, with a majority of detected cases in Honolulu County. As of the end of August 2021, all cases were attributed to the Delta variant. In early July, Honolulu County has seen a major surge in new daily cases, the largest so far, as can be seen in Fig. 7. The cases reported by the Hawai‘i department of Health climbed up to 1185 on August 29, 2021. Note that the State of Hawai‘i went from Tier 4 to Tier 5 on July 8, 2021, and despite the alarming numbers, never reversed to lower Tiers afterward.

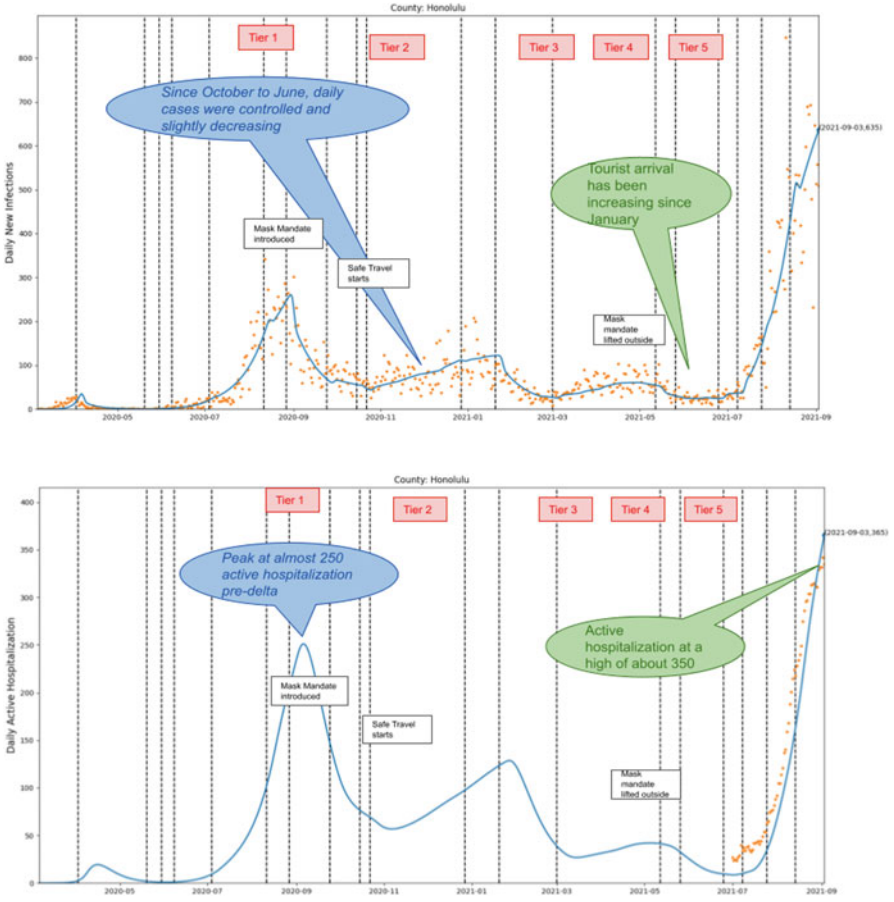


Fig. 7 Top. New Daily Cases from March 6,2020 to September 3, 2021 for Honolulu County. Right: Active Hospitalization for Honolulu County for the same period

A primary challenge in fitting the data with the model for August and September 2021 is the data variability for the new daily cases. Indeed, the observed difference in the daily case numbers from one day to the next goes up to 40%. Figure 7 bottom represents hospitalizations. Calibrating the model for hospitalizations is not trivial, the primary reason being that data was not publicly available until July 2021. Subsequently, the available data on hospitalizations came aggregated at the county level, making parameter estimation challenging since hospitalization depends on the age and health of the infected individual. However, the fit is remarkable, even with a single optimized transmission rate for those two months. Also, note that the active hospitalizations trend lags about 10 days behind the new daily cases.

The infection again spread differently in the counties after the emergence of the Delta variant. Honolulu County experienced the most rapid increase of the Delta variant, followed very closely by Hawai‘i county. Over 80% of cases due in both states were the Delta variant by mid-July. Maui was two weeks behind in reaching 80% in early August and Kaua‘i a month later in mid-August. It is unclear why these timelines are different since all counties were under the same mitigation measures and safe travel program.

The asymptomatic individuals in the CM are subdivided into 14 compartments and can develop symptoms at any stage of the infection. This allows the model to estimate the distribution of symptomatic and asymptomatic cases when fitting the model to the daily cases. Figure 8 shows the population of active asymptomatic individuals for the county of Honolulu corresponding to the model fit of the detected daily new cases coming from the department of health. For example, as of August 15, 2021, the model fit projected about 500 daily new cases and 16,587 undetected active asymptomatic individuals, among which 66% were vaccinated. The latter percentage, which is significantly larger than the previous estimate of 40% of the infected being asymptomatic, makes sense since the vaccine prevents the infected individuals from developing symptoms, and therefore many of them would go undetected. The Delta variant created a crisis for the State of Hawai‘i. The model provided 10-day forecasting each week for the department of health to inform decision-makers addressing the situation.

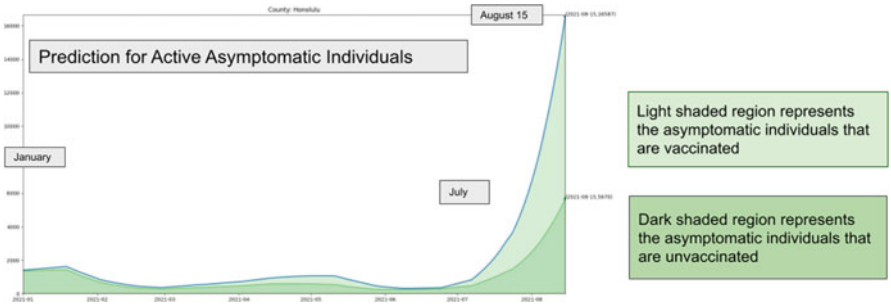


Fig. 8 Model estimate for the amount of active asymptomatic individuals necessary to fit the actual data of daily new cases

3.3 Discussion

Travelers brought COVID-19 variants into the Islands. The Delta variant emerged despite the safe travel program requiring vaccination, testing before arrival, or a period of isolation after arrival. This indicates that it might require a complete halt on travel to prevent the importation of a new infection effectively. This is, of course, not possible for many reasons, primarily economic. However, the chances are high that a delay occurs between mainland developments and the virus's arrival or the latest variant to the State of Hawai'i. This delay could be taken advantage of.

The first case of the Delta variant was identified in December 2020, well before it reached the Islands, and despite the warnings, the State of Hawai'i was unprepared to face the "Delta storm". As of September 2021, the hospitals in the State of Hawai'i are at capacity, mostly due to a shortage of health personnel. All of this happened despite the fact that prior to the Delta variant, the State-controlled the spread of the virus exceptionally well, even after reopening to travel, and despite a good vaccination rate as of early July.

4 The Pandemic Waves in New Jersey

New Jersey is a good state to explore these geospatial and temporal patterns of virus transmission for two reasons. First, New Jersey was one of the few locations in the USA with a high prevalence of COVID-19 during the first wave of the pandemic in the USA. Thus, this dataset permits exploration of how virus transmission during the first US wave compared to later waves. Second, New Jersey is the most densely populated and racially/ethnically diverse state in the USA, and thus it provides a good test case for exploring how population-level sociodemographic variables predict transmission rates over time. The following sections show spatial and temporal patterns of COVID-19 case rates across New Jersey at both the county—and municipality—level and explore how these patterns could provide insight into the variables that might be important to incorporate in future models of pandemics. Of key importance is the fact that the variables that best predict virus transmission may change with time. The variables that most strongly predict virus transmissibility in an area near the beginning of an outbreak may not be the same variables that predict transmissibility near the end of the pandemic. Understanding the temporal dynamics of critical variables can help geographically target pandemic responses to maximize health and minimize the economic impact.

4.1 Comparing New Jersey to the US

As the pandemic unfolded in New Jersey, COVID-19 case rates (number of cases, adjusted for population) varied throughout New Jersey—both across geographic areas and over time. Case rates in three regions of New Jersey (northeastern, northwestern, and southern) and the USA as a whole are shown in Fig. 9. New Jersey experienced four distinct waves of the pandemic. The first wave in New Jersey, in the spring of 2020, coincided with the virus first reaching the USA. New Jersey did not experience the wave during the summer of 2020 that reached much of the county. The second wave in New Jersey lasted from November 2020 to February 2021. The third New Jersey wave was March and April of 2021, and the fourth New Jersey wave began in July of 2021.

Predicting and mitigating pandemics requires understanding the patterns of these waves and their geographic specificity. For example, if each wave followed the same geographic pattern, it is more reasonable to model virus transmission in only one way. If, on the other hand, the geographic patterns varied depending on the wave, then a more detailed understanding is necessary.

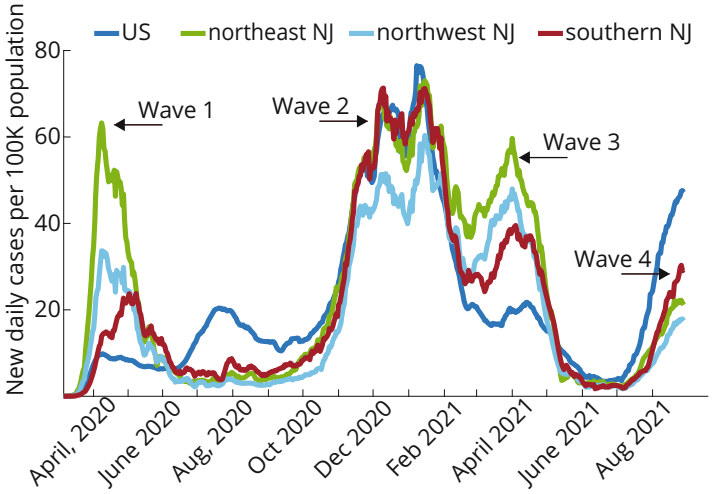


Fig. 9 The figure plots new daily cases per 100,000 residents (y-axis) as a function of time (x-axis) for four different regions: the entire USA (dark blue), northeast counties (green line: Bergen, Essex, Hudson, Middlesex, Monmouth, Ocean, Passaic, and Union Counties), northwest counties (light blue line: Hunterdon, Mercer, Morris, Somerset, Sussex, and Warren), and southern counties (red line: Atlantic, Burlington, Camden, Cape May, Cumberland, Gloucester, and Salem). For each region, the total number of cases in the region was summed and divided by the total population in the region. Data were retrieved and smoothed as described in the Methods section. The counties were divided into the three regions by initial spread of COVID-19, as defined by the number of COVID-19 cases in the county on April 1, 2020 (northeast counties: more than 1000 cases; northwest counties: between 250 and 1000 cases; southern counties: $<$250 cases per county)

4.2 *Spatial and Temporal Patterns in COVID-19 Cases in New Jersey*

The four waves of the pandemic exhibited clear regional patterns. During the first wave, case rates in New Jersey were largely predicted by proximity to New York City, the epicenter of COVID-19 in the USA. Counties in the northeast part of New Jersey, nearest to New York (green line), had the earliest and highest peak. Counties in the northwest part of the state (light blue line) had a slightly later, slightly smaller peak. And counties in southern New Jersey (furthest from New York City) had the latest, smallest peak (red line). The maps of the first wave at the county level (Fig. 10) further illustrate the broad geospatial patterns across the state during the first wave. In this map, darker shades of red indicate higher case rates during the first wave of the pandemic. The dark red areas on the map in Fig. 10 are almost exclusively located near New York City. This pattern is unsurprising and is what you might expect from travelers and commuters contracting the virus in New York City and then transmitting it to others in their home communities away from the city.

However, the regional patterns varied in later waves. During the second wave, the highest case rates occurred in the southern and northeastern parts of the state, with the northwestern parts of New Jersey having the lowest case rates (Fig. 9).

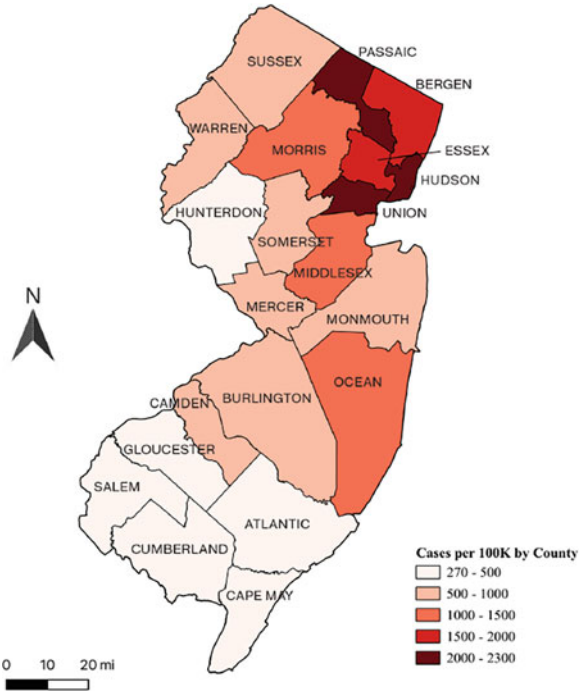


Fig. 10 The color shows the number of positive COVID-19 tests accumulated in each county between January 22, 2020 and April 27, 2020. Figure from [2] with permission from author

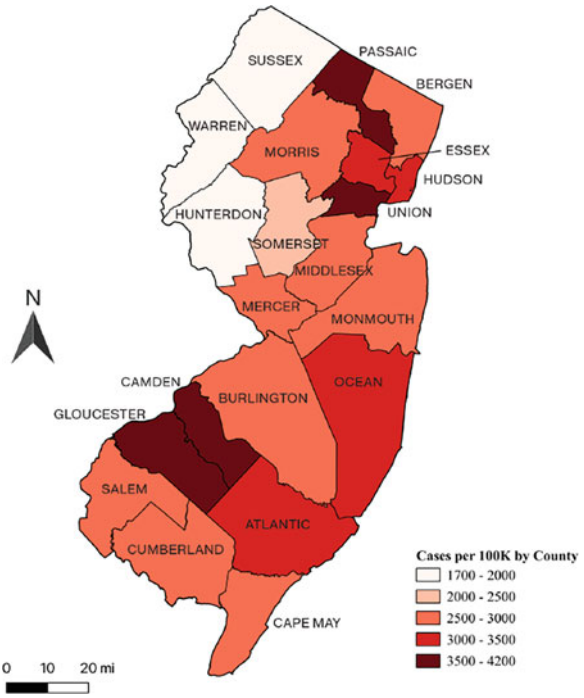


Fig. 11 The color shows the number of positive COVID-19 tests accumulated in each county between June 30, 2020 and December 13, 2020. Figure from [2] with permission from author

This pattern is further illustrated in Fig. 11, where red colors indicate the highest case rates during the second wave of the pandemic. The southern counties are much redder in the second wave than in the first wave.

The third and fourth waves were similar to the first and second. Cases in the third wave were largely predicted by proximity to New York, but as of September 2021, the southern counties of New Jersey are showing the highest case rates (Fig. 9).

4.3 Sociodemographic Variables

What can explain this pattern of results? It helps to understand the demographics of the state. Broadly speaking, the northeast part of New Jersey is very urban, whereas northwest and southern New Jersey are much less densely populated. According to data from the 2019 County Health Rankings and Roadmaps [18], the northeastern counties are 1.2% rural, the northwestern counties are 24% rural, and the southern counties are 16% rural. All of New Jersey's cities larger than 100,000 are located in the northeast. Southern New Jersey and northwest New Jersey each have only four municipalities with more than 50,000 residents, but 29 municipalities in northeast New Jersey cross this threshold.

The regions of New Jersey also differ in their socioeconomic and demographic characteristics. The densely populated northeastern New Jersey consists of a mixture of municipalities with high and low socioeconomic conditions that are close to each other. These counties are also more diverse. Counties in this region have an average median household income of about $77,000 per year and are about 55% white. Northwestern New Jersey has more consistently affluent and white suburbs. Northwestern counties have an average household income of $98,000 per year (28% **higher** than northeastern New Jersey counties) and are 72% white. In contrast, the southern region is much less affluent than both northeast and northwest New Jersey, with counties in this region averaging $68,000 per year. These counties are 66% white.

Taken together, this description of the regions of New Jersey suggests a possible explanation for the patterns across waves. Perhaps absent other mitigating factors, initial outbreaks of a virus are largely predicted by geospatial variables, while sociodemographic variables are largely irrelevant. However, after a virus has achieved wide-ranging community spread and after mitigation strategies are employed, sociodemographic variables begin to play a larger role in predicting transmission rates. For example, essential workers are more likely to come from lower socioeconomic brackets and thus have a higher risk of exposure. The first and second waves follow this pattern in New Jersey. The third wave of the virus in New Jersey followed a similar geospatial pattern as the first wave, which would be consistent with a new variant spreading throughout the region. And the fourth wave is just beginning.

4.4 Discussion

The changing patterns of COVID-19 case rates in New Jersey show clearly that the different waves of the pandemic spread across the state in different patterns. One possible explanation for this pattern is that population density and proximity to initial infection largely predict initial virus transmission, while sociodemographic variables that are correlated with behavior play a larger role in the later stages of a pandemic. However, even if an alternative hypothesis better accounts for the data, the patterns show that even with the same underlying virus, the variables best predicting virus spread can change over time. This suggests that successfully predicting transmission of novel viruses will require models to incorporate population-level variables such as sociodemographic variables, and that the relative importance of population-level variables may change over the course of a pandemic. Understanding how and when population-level sociodemographic variables influence the temporal dynamics of virus transmission is important for targeting containment and mitigation measures of future pandemics.

5 The Use of Compartmental Models in New Jersey

In the early stage of the pandemic, the model (1) was used to make predictions on the spread of the disease. More precisely, In March 2020, the Center for Computational and Integrative Biology at Rutgers University-Camden collaborated with the Senator Walter Rand Institute for Public Affairs to provide early estimates for the spread of pandemic in New Jersey, see [1]. The model was instrumental in identifying the needs for the state of New Jersey as the spread was arriving from the neighboring New York, see [1] and was cited in a letter from New Jersey Governor Phil Murphy [44]. The focus was on establishing hospital beds' needs in the eight counties of New Jersey located in the southern part, with different scenarios depending on the adopted social distancing measures. The starting point was the SEIR model proposed in [60] for COVID-19 in Wuhan, tuned using also data from the outbreak in Italy occurred in February 2020. For all the examined counties, the following were provided:

- Hospital bed availability;
- Expected peak in hospital bed demand in three scenarios;
- Expected demand exceeding capacity;
- Impact of social distancing on the basic reproduction number R_0.

The model (1) was fit to data from the Chinese Center for Disease Control and Prevention [59] and the Italian "Protezione Civile" [49]. The basic reproduction number was estimated as $R_0 = 2.68$ for the data from Wuhan and as $R_0 = 3.9$ for the Italian data. This key parameter was estimated by various papers during the pandemic in a range typically between 2 and 5. Even more challenging is to characterize the effect on R_0 of non-pharmaceutical interventions such as social distancing measures, masking, testing, and quarantining. To identify feasible values for the model parameters, we considered three zones of New Jersey, corresponding to northern, central, and southern part of the state The zones are color-coded as in Fig. 12.

5.1 Time-Evolution of the Basic Reproduction Number

In standard SEIR models, the basic reproduction number R_0 is a characteristic of the infection diseases, which is usually assumed to be constant in time. However, during the COVID-19 pandemic, such number was highly affected by non-pharmaceutical intervention and the response of population to such intervention. We used techniques developed by Thompson et al. [56], based on the *instantaneous reproduction number*. The latter is the average number of secondary cases arising from a primary case of infection [24]. In simple terms, it is an "instantaneous" transmissibility. In particular, we used the online tool available at [55]. The time between successive cases is defined as serial intervals. We used Bayesian parameter estimation to obtain

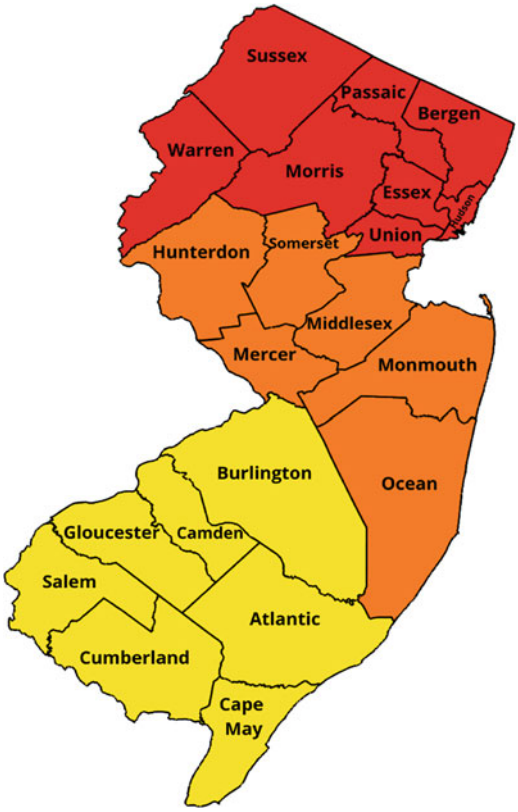


Fig. 12 Counties of New Jersey colored by zones. The red zone has a population of around 4M, the orange zone close to 3M, while the yellow one of 2M. Notice that these zones are for the regions for the New Jersey State Department of Health released hospitalization data, and are different from the sociodemographic regions in Fig. 9

posterior samples of the serial interval distribution assuming daily cases are drawn from a Poisson distribution. This allows to obtain an analytical for the posterior distribution of R_t, the time-varying reproduction number. The estimated R_t for the three regions is shown in Fig. 13 bottom right for the evolution after April 1st 2020. The time-evolution of R_t represents the impact of lockdown and other non-pharmaceutical measures.

5.2 Infected Confirmed Cases, Hospitalizations, and Deaths

Data from the Johns Hopkins public repository [22] and the New Jersey Department of Health COVID website were used to further fit the SEIR model. In particular, the repository include worldwide data for confirmed COVID-19 cases and COVID-19 related COVID-19 related deaths. COVID-19 related hospitalization data were manually copied from the New Jersey Department of Health COVID website for the relevant time period.

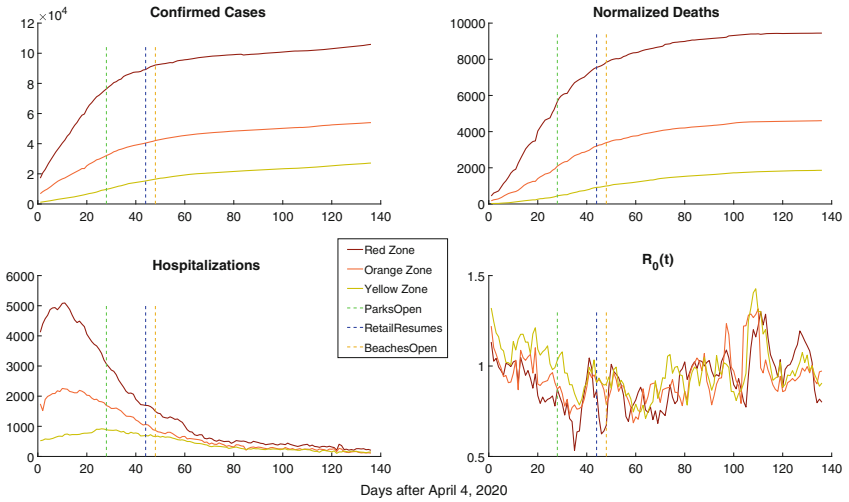


Fig. 13 Data per zone from April 1st 2020. Top Left: 7-day rolling average of confirmed cases; Top Right: 7-day rolling average of confirmed deaths. Bottom Left: Confirmed cases resulting in hospitalization; Bottom Right: Estimated reproduction rate

Due to different reporting standards during weekend days, we used seven day moving averages, which smooth out the irregularities. The data on deaths needed further treatment due to other irregularities. More precisely, on June 26, 2020 many deaths were deemed resulting from COVID-19, which previously were attributed to different cases. This was achieved by spreading such deaths weighted by the proportion of daily deaths.

Another challenge is the interpretation of confirmed cases. At the beginning of the pandemic, most people tested were in hospitals showing symptoms. However, as time progressed testing became more available, thus testing campaigns were able to detect a number of cases which show mild or no symptoms. Therefore, we considered three ways of dealing with confirmed cases:

1. All confirmed cases were symptomatic, thus in I_S.
2. All cases were symptomatic until April 20, then 50% symptomatic, thus in I_S, and 50% asymptomatic, thus in I_A by May 31.
3. All confirmed cases were symptomatic up to April 20, then the data are no more fitted.

The three cases gave rise to very similar results, thus we report results only from the third case.

We used optimization via AMPL code (“A Mathematical Programming Language”) to calibrate the model parameters to data. Data presented some discontinuities, thus we approximated the time series using degree 8 polynomial. The best parameters were determined using bounds determined by bio-medical considerations: the bounds for D_I, D_E, and D_H are shown in the estimate column

Table 2 Values of parameters per zone

Zone	max R_t	D_I	D_E	D_H	α	σ	r	E_0	I_{A_0}
Red	2.79	5.00	4.00	10.60	0.39	0.14	0.47	38,682	1.85
Orange	3.26	5.00	4.01	10.87	0.35	0.17	0.43	15,159.3	1.33136
Yellow	1.66	5.00	4.00	7.52	0.62	0.32	0.21	5141.2	407.32

of Table 2 whereas the other parameters had the following intervals: $R_0 \in [0, 15]$, $\alpha \in [0, 1]$, $\sigma \in [0, 1]$, $r \in [0, 1]$, $E_0 \in [0, 8 \times 10^6]$, $I_{A_0} \in [0, 8 \times 10^6]$.

The parameter optimization allowed to determine few characteristics of the pandemic evolution. The instantaneous reproduction number R_t showed significant differences among different zones, but these were due mainly to the initial phase of the pandemic, while data after April 1st showed a consistent pattern. The parameters of time duration of exposure, infection, and hospitalization (respectively, D_E, D_I, and D_H) showed minimal or mild variation among different zones. The incidence of asymptomatic cases ranges between 35% and 62%, lower than the initial estimate of 81% from data or early stage in Wuhan. This is in-line with data obtained later in the pandemic. Finally, the hospitalization rate ranged between 14% and 32%.

5.3 Discussion

The use of a simple CM combined with region-specific data allowed early stage studies that helped informing the decision-makers. The following data-fitting activities revealed the limitations of basic CMs. Despite augmenting the models with additional compartments to represent different stages of the disease, other challenges showed up, such the aggregated modeling of different effects (disease characteristic but also human reactions), the difficulty of interpreting data for positive, tests, the irregularity generated by the data reporting mechanism, and the dependence of model parameters on the sociodemographic characteristics of the considered regions.

6 Conclusion

We compared the evolution of the COVID-19 pandemic in Hawai‘i and New Jersey with compartmental and agent-based mathematical models. The significant geographical and sociodemographic differences in the states resulted in different transmission patterns that the models could analyze and explain. The main take-aways of our study include:

- Mathematical models can provide a framework to understand how an emerging epidemic is spreading.

- The agent-based and compartmental models have complementary advantages and disadvantages and can provide different insights into the transmission of infections.
- The accuracy of model predictions depends on the quality of the underlying data for estimating the model parameters. Reliably forecasting an emerging epidemic requires rapidly available data with sufficient granularity to estimate these parameters.
- Fitting mathematical models to data often requires information that might not be obvious to public health workers collecting data for an emerging epidemic. Early collaborations between these communities in designing the data collection and curation are needed to understand and predict future epidemics.
- The sociodemographic features of a community impact the spread of infection through that community. Reliable models must account for this heterogeneity when forecasting over widespread regions.

References

1. S.R. Allred, S.T. McQuade, N.J. Merrill, B. Piccoli, D. Spielman, C. Villacis, R. Whiting, A. Yadav, D. Zacher, D. Ziminski, Regional health system shortfalls with a novel COVID-19 model (2020) https://rand.camden.rutgers.edu/files/COVID-research-brief-with-curves03152020.pdf
2. S.R. Allred, Y. Wang, A. Yadav, R. Pletcher, T. Harig, B. Rudolph, S. Sharma, E. Ranking, S. Yan, U. Sameha, G. Arthr, S. Saxena, E. Greenfield, Municipal variation in COVID-19 case rates in New Jersey model (2021). https://rand.camden.rutgers.edu/files/COVID-research-muni-data_Brief-FINAL.pdf
3. I. Bistritz, D. Kahana, N. Bambos, I. Ben-Gal, D. Yamin, Controlling contact network topology to prevent measles outbreaks (2019), pp. 1–6
4. E. Blakemore, How measles helped destroy the Hawaiian monarchy (2019). https://www.history.com/news/hawaii-monarchy-downfall-measles-outbreak/, Retrieved Oct 2020
5. B. Bokler, Chaos and complexity in measles models: a comparative numerical study. Math. Med. Biol. J. IMA **10**(2), 83–95 (1993)
6. S. Bowong, J.J. Tewa, Mathematical analysis of a tuberculosis model with differential infectivity. Commun. Nonlinear Sci. Numer. Simul. **14**(11), 4010–4021 (2009)
7. C.C. Branas, A. Rundle, S. Pei, W. Yang, B.G. Carr, S. Sims, A. Zebrowski, R. Doorley, N. Schluger, J.W. Quinn, J. Shaman, Flattening the curve before it flattens us: hospital critical care capacity limits and mortality from novel coronavirus (SARS-CoV-2) cases in US counties. medRxiv (2020)
8. T. Britton, F. Ball, P. Trapman, A mathematical model reveals the influence of population heterogeneity on herd immunity to SARS-CoV-2. Science **369**(6505), 846–849 (2020)
9. M.G. Burch, K.A. Jacobsen, J.H. Tien, G.A. Rempala, Network-based analysis of a small Ebola outbreak (2015). Preprint, arXiv:1511.02362
10. S. Cauchemez, A. Bhattarai, T.L. Marchbanks, R.P. Fagan, S. Ostroff, N.M. Ferguson, D. Swerdlow, Pennsylvania H1N1 Working Group et al., Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza. Proc. Natl. Acad. Sci. **108**(7), 2825–2830 (2011)
11. Y.-C. Chen, P.-E. Lu, C.-S. Chang, T.-H. Liu, A time-dependent SIR model for COVID-19 with undetectable infected persons. IEEE Trans. Netw. Sci. Eng. **7**(4), 3279–3294 (2020)

12. S. Chib, E. Greenberg, Understanding the Metropolis-Hastings algorithm. Am. Stat. **49**(4), 327–335 (1995)
13. M. Chyba, A. Koniges, P. Kunwar, W. Lau, Y. Mileyko, A. Tong, COVID-19 heterogeneity in islands chain environment (2021). arXiv:2102.07646
14. M. Chyba, Y. Mileyko, O. Markovichenko, R. Carney, A. Koniges, Epidemiological model of the spread of COVID-19 in Hawaii's challenging fight against the disease, in *The Ninth International Conference on Global Health Challenges, Proceedings* (2020), pp. 32–38
15. M. Chyba, A. Koniges, P. Kunwar, W. Lau, Y. Mileyko, A. Tong, COVID-19 heterogeneity in islands chain environment (2021)
16. R.M. Colombo, M. Garavello, Well posedness and control in a nonlocal sir model. Appl. Math. Optim. **84**(1), 737–771 (2021)
17. R.M. Colombo, M. Garavello, F. Marcellini, E. Rossi, An age and space structured SIR model describing the COVID-19 pandemic. J. Math. Ind. **10**, Paper No. 22, 20 (2020)
18. County Health Rankings and Roadmaps, National data and documentation: 2010–2019 (2021) [Online]
19. A.T. Crooks, A.B. Hailegiorgis, An agent-based modeling approach applied to the spread of cholera. Environ. Model. Softw. **62**, 164–177 (2014)
20. S.C. de Greeff, H.E. de Melker, A. Westerhof, J.F.P. Schellekens, F.R. Mooi, M. van Boven, Estimation of household transmission rates of pertussis and the effect of cocooning vaccination strategies on infant pertussis. Epidemiology **23**(6), 852–860 (2012)
21. E. Demirci, A. Unal, N. Ozalp, A fractional order SEIR model with density dependent death rate. Hacet. J. Math. Stat. **40**(2), 287–295 (2011)
22. E. Dong, H. Du, L. Gardner, An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect. Dis. **20**(5), 533–534 (2020)
23. N.M. Ferguson, D. Laydon, G. Nedjati-Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunubá, G. Cuomo-Dannenburg, A. Dighe, I. Dorigatti, H. Fu, K. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, L.C. Okell, S. van Elsland, H. Thompson, R. Verity, E. Volz, H. Wang, Y. Wang, P.G.T. Walker, C. Walters, P. Winskill, C. Whittaker, C.A. Donnelly, S. Riley, A.C. Ghani, Report 9 - impact of non-pharmaceutical interventions (NPIS) to reduce COVID-19 mortality and healthcare demand (2020)
24. C. Fraser, Estimating individual and household reproduction numbers in an emerging epidemic. PLoS ONE **2**(8), 1–12 (2007)
25. M. Gatto, E. Bertuzzo, L. Mari, S. Miccoli, L. Carraro, R. Casagrandi, A. Rinaldo, Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures. Proc. Natl. Acad. Sci. **117**(9), 10484–10491 (2020)
26. G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A. Di Filippo, A. Di Matteo, M. Colaneri, Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nat. Med. **26**(6), 855–860 (2020)
27. C. Groendyke, D. Welch, D.R. Hunter, A network-based analysis of the 1861 Hagelloch measles data. Biometrics **68**(3), 755–765 (2012)
28. Hawaii Population Model, Hawai'i data collaborative (2021). https://www.hawaiidata.org/hawaii-population-model
29. E.M.T. Hendrix, G. Boglárka et al., *Introduction to Nonlinear and Global Optimization*, vol. 37 (Springer, Berlin, 2010)
30. D. Herman, Shutting down Hawai'i: a historical perspective on epidemics in the Islands (2020). https://www.smithsonianmag.com/history/shutting-down-hawaii-historical-perspective/-epidemics-islands-180974506/, retrieved Oct 2020
31. H.W. Hethcote, The mathematics of infectious diseases. SIAM Rev. **42**(4), 599–653 (2000)
32. Z. Jin, J. Zhang, L.-P. Song, G.-Q. Sun, J. Kan, H. Zhu, Modelling and analysis of influenza A (H1N1) on networks. BMC Publ. Health **11**(1), 1–9 (2011)
33. V. Kala, K. Guo, E. Swantek, A. Tong, M. Chyba, Y. Mileyko, C. Gray, T. Lee, A.E. Koniges, Pandemics in Hawaii: 1918 influenza and COVID-19, in *The Ninth International Conference on Global Health Challenges GLOBAL HEALTH 2020*, IARIA (2020)

34. W.O. Kermack, A.G. McKendrick, Contributions to the mathematical theory of epidemics. II. The problem of endemicity. Proc. R. Soc. Lond. A Contain. Pap. Math. Phys. Character **138**(834), 55–83 (1932)
35. C.C. Kerr, R.M. Stuart, D. Mistry, R.G. Abeysuriya, G. Hart, K. Rosenfeld, P. Selvaraj, R.C. Núñez, B. Hagedorn, L. George, A. Izzo, A. Palmer, D. Delport, C. Bennette, B. Wagner, S. Chang, J.A. Cohen, J. Panovska-Griffiths, M. Jastrzębski, A.P. Oron, E. Wenger, M. Famulare, D.J. Klein, Covasim: an agent-based model of COVID-19 dynamics and interventions. medRxiv (2020)
36. D. Kim, A. Quaini, Coupling kinetic theory approaches for pedestrian dynamics and disease contagion in a confined environment. Math. Models Methods Appl. Sci. **30**(10), 1893–1915 (2020)
37. P. Kunwar, O. Markovichenko, M. Chyba, Y. Mileyko, A. Koniges, T. Lee, A study of computational and conceptual complexities of compartment and agent based models (2021). arXiv:2108.11546
38. P.E. Lekone, B.F. Finkenstädt, Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study. Biometrics **62**(4), 1170–1177 (2006)
39. J.O. Lloyd-Smith, A.P. Galvani, W.M. Getz, Curtailing transmission of severe acute respiratory syndrome within a community and its hospital. Proc. Biol. Sci. **270**, 1979–1989 (2003)
40. D.K. Mamo, P.R. Koya, Mathematical modeling and simulation study of SEIR disease and data fitting of Ebola epidemic spreading in West Africa. J. Multidiscip. Eng. Sci. Technol. **2**(1), 106–114 (2015)
41. S. Mei, D. van de Vijver, L. Xuan, Y. Zhu, P.M.A. Sloot, Quantitatively evaluating interventions in the influenza A (H1N1) epidemic on China campus grounded on individual-based simulations. Procedia Comput. Sci. **1**(1), 1675–1682 (2010)
42. C.J.E. Metcalf, D.H. Morris, S.W. Park, Mathematical models to guide pandemic response. Science **369**(6502), 368–369 (2020)
43. K.R. Moran, G. Fairchild, N. Generous, K. Hickmann, D. Osthus, R. Priedhorsky, J. Hyman, S.Y. Del Valle. Epidemic forecasting is messier than weather forecasting: the role of human behavior and internet data streams in epidemic forecast. J. Infect. Dis. **214**(suppl. 4), S404–S408 (2016)
44. P.D. Murphy, Letter to the President Donald J. Trump, 17 March 2020. http://d31hzlhk6di2h5.cloudfront.net/20200317/3c/e6/ea/5b/71a343b469cf7732d3a12e0e/President_Trump_Ltr_re_COVID19_3.17.20.pdf
45. N. Özalp, E. Demirci, A fractional order SEIR model with vertical transmission. Math. Comput. Model. **54**(1–2), 1–6 (2011)
46. C.W. Reynolds, Flocks, herds and schools: a distributed behavioral model, in *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques* (1987), pp. 25–34
47. A. Rizzo, B. Pedalino, M. Porfiri, A network model for Ebola spreading. J. Theor. Biol. **394**, 212–222 (2016)
48. P. Rohani, X. Zhong, A.A. King, Contact network structure explains the changing epidemiology of pertussis. Science **330**(6006), 982–985 (2010)
49. U. Rosini, March update: data provided by the ministry of health and acquired by the regions in the operating board (2020). github:https://github.com/pcm-dpc/COVID-19/blob/master/schede-riepilogative/regioni/dpc-covid19-ita-scheda-regioni-20200302.pdf
50. N.W. Ruktanonchai, J.R. Floyd, S. Lai, C.W. Ruktanonchai, A. Sadilek, P. Rente-Lourenco, X. Ben, A. Carioli, J. Gwinn, J.E. Steele et al., Assessing the impact of coordinated COVID-19 exit strategies across europe. Science **369**(6510), 1465–1470 (2020)
51. T.C. Schelling, Dynamic models of segregation. J. Math. Sociol. **1**(2), 143–186 (1971)
52. T.C. Schelling, *Micromotives and Macrobehavior* (WW Norton, New York, 1978)
53. R. Schmitt, E.C. Nordyke, Death in Hawai'i: the epidemics of 1848–1849. Hawaii J. Hist. **35**, 1–13 (2001)
54. S. Side, W. Sanusi, M.K. Aidid, S. Sidjara, Global stability of SIR and SEIR model for Tuberculosis disease transmission with Lyapunov function method. Asian J. Appl. Sci. **9**(3), 87–96 (2016)

55. R.N. Thompson, J.E. Stockwin, R.D. van Gaalen, J.A. Polonsky, Z.N. Kamvar, P.A. Demarsh, E. Dahlqwist, S. Li, E. Miguel, T. Jombart et al., Epiestim app (2019). https://shiny.dide.imperial.ac.uk/epiestim
56. R.N. Thompson, J.E. Stockwin, R.D. van Gaalen, J.A. Polonsky, Z.N. Kamvar, P.A. Demarsh, E. Dahlqwist, S. Li, E. Miguel, T. Jombart et al., Improved inference of time-varying reproduction numbers during infectious disease outbreaks. Epidemics **29**, 100356 (2019)
57. Travel Data, COVID-19, state of Hawaii portal. https://hawaiicovid19.com/travel/data/
58. A. Vespignani, H. Tian, C. Dye, J.O. Lloyd-Smith, R.M. Eggo, M. Shrestha, S.V. Scarpino, B. Gutierrez, M.U.G. Kraemer, J. Wu et al., Modelling COVID-19. Nat. Rev. Phys. **2**(6), 279–281 (2020)
59. Z. Wu, J.M. McGoogan, Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72,314 cases from the Chinese center for disease control and prevention. JAMA **323**(13), 1239–1242 (2020)
60. J.T. Wu, K. Leung, G.M. Leung, Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet **395**(10225), 689–697 (2020)
61. J. Zhang, M. Litvinova, Y. Liang, Y. Wang, W. Wang, S. Zhao, Q. Wu, S. Merler, C. Viboud, A. Vespignani et al., Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. Science **368**(6498), 1481–1486 (2020)
62. J. Zhang, M. Litvinova, W. Wang, Y. Wang, X. Deng, X. Chen, M. Li, W. Zheng, L. Yi, X. Chen et al., Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. Lancet Infect. Dis. **20**(7), 793–802 (2020)
63. F. Zhou, T. Yu, R. Du, G. Fan, Y. Liu, Z. Liu, J. Xiang, Y. Wang, B. Song, X. Gu et al., Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet **395**(10229), 1054–1062 (2020)

A Novel Point Process Model for COVID-19: Multivariate Recursive Hawkes Process

Bohan Chen, Pujan Shrestha, Andrea L. Bertozzi, George Mohler, and Frederic Schoenberg

1 Introduction

1.1 Hawkes Point Process Modeling of Infectious Diseases

Recent research has been devoted to modeling infectious disease transmission using branching point processes such as the Hawkes process [19, 26, 29]. In the early weeks of the COVID-19 pandemic, Bertozzi et al. [3] used a Hawkes process model with dynamic reproduction number to model regional scale transmission. In [8], mobility indices and spatial covariates are incorporated in Hawkes process models of COVID-19 to improve forecasts. Hawkes processes are also used in [20] to model cases in France, where the authors simulate two scenarios, with and without lockdown. Their experimental results shed some light on how shutting down a city can contain the spread of the disease. Following the idea of [10], a two-phase dynamic contagion process is introduced in [7] to model COVID-19 before and after

B. Chen · A. L. Bertozzi (✉)
Department of Mathematics, University of California, Los Angeles, Los Angeles, CA, USA
e-mail: bhchenyz@g.ucla.edu; bertozzi@ucla.edu

P. Shrestha
Department of Mathematics, Louisiana State University, Baton Rouge, LA, USA
e-mail: pshres6@lsu.edu

G. Mohler
Department of Computer Science, Boston College, Boston, MA, USA
e-mail: mohlerg@bc.edu

F. Schoenberg
Department of Statistics, University of California, Los Angeles, Los Angeles, CA, USA
e-mail: frederic@stat.ucla.edu

N. Bellomo, M. A. J. Chaplain (eds.), *Predicting Pandemics in a Globally Connected World, Volume 1*, Modeling and Simulation in Science, Engineering and Technology, https://doi.org/10.1007/978-3-030-96562-4_5

preventative actions. The underlying idea is to use two kinds of point processes to model COVID-19 cases. Before the lock-down, exogenous cases are modeled with externally excited triggering functions, while after the lockdown a monotonically decaying function is adopted to model the intensity of the events triggered by exogenous cases.

Compared to compartmental models that consider counts of events over discrete-time intervals, point processes are continuous-time, stochastic models of infectious disease transmission that treat events, and their interactions, individually. In particular, the number of confirmed cases (or deaths) is modeled as a counting process, $\{N(t), t \geq 0\}$ with values that are non-negative, integer, and non-decreasing. In other words, for $s < t$, $N(t) - N(s)$ corresponds to the number of events occurring during the interval $(s, t]$. The conditional intensity, $\lambda(t|\mathcal{H}_t)$, is used to characterize the evolution of a counting process. This conditional intensity is defined by:

$$\lambda(t|\mathcal{H}_t) = \lim_{\Delta \to 0} \frac{1}{\Delta t} \mathrm{E}(N(t + \Delta t) - N(t)|\mathcal{H}_t), \tag{1}$$

where $\mathcal{H}_t$ is a filtration such that $\mathcal{H}_t$ contains all historical information on points occurring before time t.

A Hawkes process [16, 17], also known as a self-exciting process, is a special kind of point process in which background events occur according to a Poisson process, and then these background events trigger other events (e.g., secondary infection, earthquake aftershock), which in turn may trigger other events (and so on). The conditional intensity of a Hawkes process is given by:

$$\lambda(t) = \mu + K \int_0^{t-} g(t - s) dN(s) = \mu + K \sum_{t_i < t} g(t - t_i), \tag{2}$$

where $\{t_i$ is the time of occurrence of the ith event of the process and $g : \mathbb{R}^+ \to \mathbb{R}^+$ is a kernel density-"triggering" function that expresses the influence of past events $\{t_i\}$ on the current intensity $\lambda(t)$. In Eq. (2), K is the productivity (reproduction number) indicating the expected number of offspring directly triggered by each event, and μ is the background rate representing determining the intensity of spontaneous (Poisson) background events.

1.2 Multivariate Hawkes Processes

In situations where events occur on a network, for example, a viral tweet and subsequent retweets, a multivariate Hawkes process can be used to model event cascades across nodes of the network. The multivariate Hawkes process [21] is an extension of the univariate Hawkes process in Eq. (2) when there are multiple sub-processes indexed by $u = 1, 2, \ldots, U$. The conditional intensity $\lambda_u(t)$ corresponding to the sub-process u can be expressed as:

$$\lambda_u(t) = \mu_u + \sum_{j=1}^{U} K_{j,u} \int_0^{t_-} g_j(t-s) dN(s) = \mu_u + \sum_{j=1}^{U} \sum_{i:t_i<t} K_{j,u} g_i(t-t_i), \tag{3}$$

where event j has both a time, t_j, and a mark (network node index), u_j. Each directional edge has its own productivity number $K_{j,u}$, which gives the expected number of events of index u that are triggered by one event of index j. In Eq. (3), g_u is the time-triggering function of node u and μ_u is the background rate of node u.

The multivariate Hawkes process considers both self-exciting effects (within a node of the network), along with mutual excitement across the edges of a network. The events from one sub-process can contribute to the intensity function of another sub-process through the kernel. This multivariate model has been used in network reconstruction and modeling dynamics across heterogeneous groups due [31]. In [13], for instance, a temporal multivariate Hawkes process is proposed for inferring leadership from an email network.

While some work has been dedicated to monitoring the progress of diseases only based on the number of daily reported cases, many have also tried to tackle the problem through spatiotemporal approaches. For a two-dimensional multivariate spatiotemporal model, events of the point process are expressed by $\{(x_i, y_i, t_i, u_i\}_{i=1}^{N}$, where x_i, y_i are spatial coordinates, t_i is the time, and u_i is the index of the sub-process. In this case the intensity function includes the spatial coordinates:

$$\begin{aligned} \lambda_u(x, y, t) &= \mu_u(x, y) + \sum_{j=1}^{U} K_{j,u} \int_0^{t_-} g_u(d_k, t-s) dN(s) \\ &= \mu_u(x, y) + \sum_{j=1}^{U} K_{j,u} \sum_{i:t_i<t} g_j(d_i, t-t_i), \end{aligned} \tag{4}$$

where $d_i = \sqrt{(x-x_i)^2 + (y-y_i)^2}$ and g_u models how triggering decays over a distance, d_i. In [12, 24, 25], spatiotemporal point processes are introduced to model how the geo-locations would affect the triggering functions. Gaussian kernels are used to model the decay behaviors with respect to distances. In [24] power-law and Gaussian functions are compared, and experimental results demonstrate better performance on the power-law function than naive Gaussian. In [33], the authors developed a fast estimation method of the parameters of the spatiotemporal multivariate Hawkes model and implemented it on the crime-topic networks in Los Angeles.

1.3 Recursive Hawkes Processes

One drawback of the univariate and multivariate Hawkes processes above is that the productivity, or reproduction number, is fixed across time. However, in the case of disease transmission, public health measures, social distancing, or other factors may reduce the reproduction number. To deal with these issues, a "recursive" point process model is introduced in [28] to study the spread of contagious diseases when the reproduction number depends on the conditional intensity, expressed by:

$$\lambda(t) = \mu + \int_0^{t-} h(\lambda(s))g(t-s)dN(s) = \mu + \sum_{t_j<t} h(\lambda(t_j))g(t-t_j). \tag{5}$$

Compared with the univariate Hawkes process in Eq. (2), the recursive model has an extra function h, called the recursive function, which is assumed to be a non-increasing function of the intensity. The recursive Hawkes model can be used for novel viruses when the disease has a high transmission rate at the start of an outbreak, due to the high prevalence of susceptible people and a lack of public health interventions, and then later has a lower transmission rate, as more people are infected and recover from the disease, and as policies and behaviors change to mitigate the spread of the disease. The recursive model can also be used to evaluate the effect of imposed interventions during a pandemic [28, 32].

Schoenberg et al. [28] applied the recursive model to fit Rocky Mountain Spotted Fever cases in California from 1960–2011, and the results suggest that the recursive model outperforms the traditional Hawkes process in terms of goodness-of-fit. Kaplan et al. [18] took the model a step further by estimating the recursive function defining the productivity non-parametrically, rather than being characterized by a parametric decay function. Geo-location was also included in that model, through an exponential decay in space in the triggering function with respect to the distance between a primary and a secondary infection.

1.4 Outline

In the present chapter, the goal is to extend the recursive Hawkes process to the multivariate setting. The multivariate recursive Hawkes process has the advantages of the multivariate Hawkes process, which can capture disease transmission across a network, while also adapting to changes in reproduction through public health interventions, social distancing, natural immunity, and vaccination. The outline of the chapter is as follows. Section 2 includes the introduction of the proposed model along with several theoretical results on existence, expectation, and variance. Section 3 covers parameter fitting algorithms and simulation methods for the multivariate recursive Hawkes process. Section 4 concerns the reconstruction of multivariate point processes from data with imprecise times, and Sect. 5 presents

results from the multivariate recursive Hawkes process when the model is applied to synthetic data and Covid-19 case data from California.

2 Theoretical Properties of Temporal Multivariate Recursive Hawkes Models

For the model, it is assumed that the productivity at a site is nonzero, has variability in the sense that it dynamically may change over time, and is recursively related to the conditional intensity at previous event times for all sites. At the start of an outbreak in a new location, the conditional intensity λ is initially low. However, the rate of cases will increase exponentially (when the productivity is above 1) in the absence of public health interventions by local government or natural immunity in the population (due to past infection). During later stages of the outbreak, public health interventions, changes in behavior (e.g., social distancing), vaccinations, and natural immunity can result in a lower rate of transmission despite a high conditional intensity. These assumptions align well with the contemporary literature on outbreak modeling [22] and disease epidemics [6], and the presented model aims to capture such phenomena. The univariate recursion case was explored by Schoenberg et al. in [28]. In the present work, the intensity formula (5) is expanded to the multivariate case. The model is considered via a U-tuple conditional intensity, $\lambda(t) = [\lambda_u(t)]_{u=1}^{U}$, where for each node u (corresponding to a particular location, e.g., city, county state, depending on the resolution of the model)

$$\lambda_u(t) = \mu_u + \sum_{j=1}^{U} \int_0^{t_-} H(\lambda_j(t'), j, u) g(t - t') dN_{t'} \tag{6}$$

Assumptions:

- Background rate: $\mu = [\mu_u]_{u=1}^{U}$ with $\mu_u > 0 \; \forall \, u$
- Triggering function: $g > 0$ is a probability density function
- Productivity function: $H(\lambda_j(t'), u, j) = K_{j,u} h(\lambda_j(t'))$ where $K_{j,u}$ describes the productivity for a new event in sub-process u being caused by sub-process j and $h(\lambda_j(t))$ is the recursive function. $K_{j,u}$ is assumed to be bounded

From the above discussion, it is inferred that H should generally be decreasing. When restricting the number of sub-processes to a single one and taking $h(x) = x^{-1}$, then the model reduces to the standard recursive case proposed by Schoenberg et al. [28]. Following sections include the proof of the existence of a point process with the given conditional intensity function (6) and the derivation for formulas of the mean and variance. The results vary for different choices of the recursive function. Work here is motivated by the univariate case in [28].

2.1 Existence

Proposition 1 *Given a complete probability space, a counting process, N_t, with conditional intensity satisfying (6) can be constructed with $H(x,\cdot,\cdot)$ bounded, non-negative and decreasing in x.*

Proof Let $(e_k)_{k\geq 1}$ be a sequence of independent exponential random variables with unit mean. Set $T_0 = 0$ and for each u, define a sequence of stopping times, $\{T_k^u\}$ by

$$T_{k+1}^u = \inf_t \left\{ t > T_k^u : \int_{T_k^u}^{t^-} (\mu_u + \sum_{j=1}^{U} \int_0^t H(\lambda_j(T_i), j, u) g(s - T_i) dN_{t'}^u) ds \geq e_{k+1} \right\}.$$

Define, for $k \geq 1$ the sequence of sub-processes $N_t^{u,(k)} = \sum_{i=1}^{k} \mathbb{I}_{\{T_i^u \leq t\}}$, and define $N_t^{(k)} = \left[N_t^{u,(k)} \right]_{u=1}^{U}$. Note that, $(x_1, x_2, \cdots, x_n) \leq (y_1, y_2, \cdots, y_n)$ if and only if each $x_j \leq y_j$. Using the Random Time Change Theorem [9], $N_t^{(k)}$ is a counting process with conditional intensity $\lambda^k(t) = \left[\lambda_u^k(t) \right]_{u=1}^{U}$ specified by

$$\lambda_u^k(t) = \begin{cases} \mu_u + \sum_{j=1}^{U} \int_0^{t^-} H(\lambda_j(v), j, u) g(t - v) dN_v^{(k)} & \text{if } N_{t_-}^{(k)} < k, \\ 0 & \text{if } N_{t_-}^{(k)} = k, \end{cases}$$

where $N_{t^-}^{(k)} = \lim_{v \nearrow t} N_v^{(k)}$. Then use truncated counting processes to make estimations for the limit case. Now, let $N_t = \lim_{k\to\infty} N_t^{(k)}$. It should be focused on showing that the model is well defined and does not include a finite time blowup. For $k \geq 1$,

$$\begin{aligned} \mathbb{E}\left[N_t^{(k)} \right] &= \mathbb{E}\left[\int_0^{t_-} \lambda^k(s) ds \right] \\ &= \mathbb{E}\left[\int_0^{t_-} \left[\lambda_u^k(s) \right]_{u=1}^{U} ds \right] = \left[\mathbb{E}\left[\int_0^{t_-} \lambda_u^k(s) ds \right] \right]_{u=1}^{U}. \end{aligned} \tag{7}$$

By assumption, U is a finite number, which gives the reason to focus on arbitrary u. Notice the fact that productivity function H is bounded by a finite number, denoted as M. Successively, apply a convolution property for point processes (Lemma 22 in [11]) and Fubini's Theorem to obtain

$$\mathbb{E}\left[\int_0^{t_-} \lambda_u^k(s)ds\right] \leq \mu_u t + M\mathbb{E}\left[\int_0^{t_-} \sum_{j=1}^{U} \int_0^{s_-} g(t-t')dN_{t'}^{(k)}ds\right]$$
$$= \mu_u t + MU\mathbb{E}\left[\int_0^{t_-} g(t-s)N_s^{(k)}ds\right]$$
$$= \mu_u t + MU\int_0^{t_-} g(t-s)\mathbb{E}\left[N_s^{(k)}\right]ds,$$

Combine the results for a single node u in the U-tuple equation (7) to get

$$\mathbb{E}\left[N_t^{(k)}\right] \leq \left[\mathbb{E}\left[\int_0^{t_-} \lambda_u^k(s)ds\right]\right]_{u=1}^{U} \leq \mu t + MU\int_0^{t_-} g(t-s)\mathbb{E}\left[N_s^{(k)}\right]ds.$$

Then, via Gronwall's inequality, $\sup_k \mathbb{E}\left[N_t^{(k)}\right] \leq \mu t\kappa$, where κ is a constant depending on M, U, g, t and provided g is locally integrable. This upper bound is uniform for all $k > 0$ and a fixed t. By the Monotone Convergence Theorem, it is obtained that $\mathbb{E}[N_t] < \infty$. This implies that $N_t < \infty$ almost surely. From this, the conditional intensity of N_t satisfies equation (6). □

By existence, it guaranteed that a probability space can be found where the point process does not blow up in finite time. The uniqueness in distribution of the limit is well supported by contemporary literature (Theorem 6.55 in [21]).

2.2 Mean

The expectation of $N(t)$ being finite allows us to exploit the Fubini theorem, the martingale property of point processes, and Campbell's Theorem [34] to find the mean of the multivariate recursive process with conditional intensity defined by (6).

Proposition 2 *Define* $\zeta = [0, T]$. *Under the assumption that*

$$\lim_{T\to\infty} \int_0^{T-t'} g(t)\,dt = 1, \quad \forall t', \tag{8}$$

there is an estimation of the mean as

$$\lim_{T\to\infty} \frac{1}{T}\mathbb{E}[N(\zeta)] = \mu + \frac{1}{T}\left[\sum_{j=1}^{U} \mathbb{E}\left(\int_0^T H(\lambda_j(t'), j, u)\lambda_u(t')dt'\right)\right]_{u=1}^{U}. \tag{9}$$

Proof

$$\frac{1}{T}\mathbb{E}[N(\zeta)] = \frac{1}{T}\mathbb{E}\left[\int_{\zeta} dN_t\right] = \frac{1}{T}\mathbb{E}\left[\int_0^T \lambda(t)dt\right] = \frac{1}{T}\left[\int_0^T \mathbb{E}\left[\lambda_u(t)\right]dt\right]_{u=1}^{U}$$

$$= \frac{1}{T}\left[\mathbb{E}\left(\int_0^T \left(\mu_u + \int_0^t \sum_{j=1}^{U} H(\lambda_j(t'), j, u)g(t-t')dN_{t'}^u\right)dt\right)\right]_{u=1}^{U}$$

$$= [\mu_u]_{u=1}^{U} + \frac{1}{T}\left[\int_0^T \mathbb{E}\left(\mu_u + \int_0^t \sum_{j=1}^{U} H(\lambda_j(t'), j, u)g(t-t')dN_{t'}^u\right)dt\right]_{u=1}^{U}$$

$$= \mu + \frac{1}{T}\left[\int_0^T \sum_{j=1}^{U} \mathbb{E}\left(\int_0^t H(\lambda_j(t'), j, u)g(t-t')dN_{t'}^u\right)dt\right]_{u=1}^{U}$$

$$= \mu + \frac{1}{T}\left[\sum_{j=1}^{U} \mathbb{E}\left(\int_0^T \int_0^t H(\lambda_j(t'), j, u)\lambda_u(t')g(t-t')dt'dt\right)\right]_{u=1}^{U}$$

$$= \mu + \frac{1}{T}\left[\sum_{j=1}^{U} \mathbb{E}\left(\int_0^T H(\lambda_j(t'), j, u)\lambda_u(t')\left(\int_0^{T-t'} g(t)dt\right)dt'\right)\right]_{u=1}^{U}$$

$$\longrightarrow \mu + \frac{1}{T}\left[\sum_{j=1}^{U} \mathbb{E}\left(\int_0^T H(\lambda_j(t'), j, u)\lambda_u(t')dt'\right)\right]_{u=1}^{U}, \ T \to \infty$$

$$= \mu + \frac{1}{T}\left[\sum_{j=1}^{U} \mathbb{E}\left(\int_0^T K_{j,u}h(\lambda_j(t'))\lambda_u(t')dt'\right)\right]_{u=1}^{U}.$$

□

Remark 1 If the assumption (8) is not true, then equation (9) is an approximation. Without such an assumptions, the impact on consistent estimation of parameters for univariate recursive model is investigated in [27]. The motivation for the analysis into asymptotic mean is to check the stability of the model for large T. The investigation provides an insight into the importance of the recursive function h, while this result does not provide insight into h itself.

2.3 Variance

With the aim of classifying the class of "nice" functions for h, the univariate case in [28] is used for inspiration.

Proposition 3 *According to proposition 2, under an assumption about the time triggering function g, there is an estimation (9) of the mean of the point process. Based on this estimation, under two assumptions of the recursive function h:*

1. *Define* $M_{j,u}(t) = \int_0^t h\left(\lambda_j(t')\right) dN_{t'}^u$. *assume that for each* $\{j, u\}$,*there exists an increasing deterministic function,* $\alpha_t^{j,u}$, *such that* $M_{j,u}(t) - \alpha_t^{j,u}$ *is martingale;*
2. *For each* u, $j \in U$, $\int_0^T g(T-t)\alpha_t^{j,u} dt < \infty$;

there is an estimation of the variance:

$$\sup_T \left\{ T\mathbb{E}\left[\left(T^{-1}N(T) - \left(\mu + \frac{1}{T}\left[\sum_{j=1}^{U} \mathbb{E}\left[\int_0^T K_{j,u}\lambda_u(t) h\left(\lambda_j(t)\right) dt \right] \right]_{u=1}^{U} \right) \right)^2 \right] \right\}. \tag{10}$$

It will be shown that the estimation (10) is always finite.

Proof In order to obtain these bounds, rewrite the inner term in (10) by making the following modifications,

$$\begin{aligned} & T^{-1}N(T) - \left(\mu + \frac{1}{T}\left[\sum_{j=1}^{U} \mathbb{E}\left[\int_0^T K_{j,u}\lambda_u(t) h\left(\lambda_j(t)\right) dt \right] \right]_{u=1}^{U} \right) \\ =& \frac{1}{T}\left(N(T) - \int_0^T \lambda(t) dt \right) \\ & - \left(\mu + \frac{1}{T}\left[\sum_{j=1}^{U} \mathbb{E}\left[\int_0^T K_{j,u}\lambda_u(t) h\left(\lambda_j(t)\right) dt \right] \right]_{u=1}^{U} - \frac{1}{T}\int_0^T \lambda(t) dt \right). \end{aligned}$$

For the ease of notation, define

$$A_T = \frac{1}{T}\left(N(T) - \int_0^T \lambda(t) dt \right)$$

$$B_T = \mu + \frac{1}{T}\left[\sum_{j=1}^{U} \mathbb{E}\left[\int_0^T K_{j,u}\lambda_u(t) h\left(\lambda_j(t)\right) dt \right] \right]_{u=1}^{U} - \frac{1}{T}\int_0^T \lambda(t) dt$$

Since the inner term in (10) equals $A_T - B_T$, use the inequality $(a-b)^2 \leq 2(a^2+b^2)$ to majorize. The goal is to show $T\mathbb{E}\left[{A_T}^2\right]$ and $T\mathbb{E}\left[{B_T}^2\right]$ are both upper bounded conditional on the selection of the recursive function h. Consider focus on the first term, A_T. Refer to [4] for more information on the martingale properties of point processes. Using the fact that $N_T - \int_0^T \lambda(t)dt$ is a martingale with the compensator $\left\langle N_\bullet - \int_0^\bullet \lambda(t)dt \right\rangle_T = \int_0^T \lambda(t)dt$, it comes to

$$\begin{aligned} T\mathbb{E}\left[{A_T}^2\right] &= T\mathbb{E}\left[\left(\frac{1}{T}(N_T - \int_0^T \lambda(t)dt)\right)^2\right] \\ &= \frac{1}{T}\mathbb{E}\left[\langle N_\bullet - \int_0^\bullet \lambda(t)dt\rangle_T\right] = \frac{1}{T}\int_0^T \mathbb{E}[\lambda(t)]dt \end{aligned}$$

Using the same argument flow as in Proposition 1 but instead now for $\lambda(t)$, it shows that $\sup_t \mathbb{E}[\lambda(t)]$ is finite. Thus,

$$\sup_T T\mathbb{E}[A_t^2] < \infty. \tag{11}$$

Now let us turn focus onto B_T.

$$\begin{aligned} B_T &= \frac{1}{T}\int_0^T \lambda(t)dt - \left(\mu + \frac{1}{T}\left[\sum_{j=1}^U \mathbb{E}\left[\int_0^T K_{j,u}\lambda_u(t)h\left(\lambda_j(t)\right)dt\right]\right]_{u=1}^U\right) \\ &= \frac{1}{T}\int_0^T [\lambda_u(t)]_{u=1}^U dt - \left(\mu + \frac{1}{T}\left[\sum_{j=1}^U \mathbb{E}\left[\int_0^T K_{j,u}\lambda_u(t)h\left(\lambda_j(t)\right)dt\right]\right]_{u=1}^U\right) \\ &= \frac{1}{T}\int_0^T [\lambda_u(t) - \mu_u]_{u=1}^U dt \\ &\quad - \frac{1}{T}\left[\sum_{j=1}^U \mathbb{E}\left[\int_0^T K_{j,u}\lambda_u(t)h\left(\lambda_j(t)\right)dt\right]\right]_{u=1}^U \\ &= \frac{1}{T}\left[\int_0^T \sum_{j=1}^U \int_0^{t_-} K_{j,u}h(\lambda_j(t'))g(t-t')dN_{t'}^u dt\right]_{u=1}^U \\ &\quad - \frac{1}{T}\left[\sum_{j=1}^U \mathbb{E}\left[\int_0^T K_{j,u}\lambda_u(t)h\left(\lambda_j(t)\right)dt\right]\right]_{u=1}^U \end{aligned}$$

Now for any node u, partition the interval $[0, T]$, by the event arrival times for the u-th process. Here, $0 = t_0 < t_1 < t_2 < \cdots < t_n < T$. The summation definition of the conditional intensity function for Hawkes Process is used to simplify the first term.

$$\begin{aligned}
&\int_0^T \left(\sum_{j=1}^{U} \int_0^{t_-} K_{j,u} h(\lambda_j(t')) g(t - t') dN_{t'}^u \right) dt \\
&= \sum_{j=1}^{U} K_{j,u} \int_0^T \sum_{t'<t} h(\lambda_j(t')) g(t - t') dt \\
&= \sum_{j=1}^{U} K_{j,u} \sum_{i=1}^{U} \int_{t_i}^{t_i+1} \sum_{k=1}^{i} h\left(\lambda_j(t_k)\right) g(t - t_k) dt \\
&= \sum_{j=1}^{U} K_{j,u} \sum_{k=1}^{U} \sum_{i=k}^{U} h\left(\lambda_j(t_k)\right) \int_{t_i}^{t_i+1} g(t - t_k) dt \\
&= \sum_{j=1}^{U} K_{j,u} \sum_{k=1}^{U} h\left(\lambda_j(t_k)\right) \int_{t_i}^{T} g(t - t_k) dt \\
&= \sum_{j=1}^{U} K_{j,u} \int_0^T h\left(\lambda_j(t')\right) \int_{t'}^{T} g(t - t') dt dN_{t'}^u \\
&= \sum_{j=1}^{U} K_{j,u} \int_0^T g(T - t) \int_0^t h\left(\lambda_j(t')\right) dN_{t'}^u dt
\end{aligned}$$

Define $M_{j,u}(t) = \int_0^t h\left(\lambda_j(t')\right) dN_{t'}^u$. According to the assumption 1, for each $\{j, u\}$, there exists an increasing deterministic function, $\alpha_t^{j,u}$, such that $M_{j,u}(t) - \alpha_t^{j,u}$ is martingale. Writing this term in the expression for B_T, it is the equality

$$\begin{aligned}
B_T = &\frac{1}{T} \left[\sum_{j=1}^{U} K_{j,u} \int_0^T g(T - t) M_{j,u}(t) dt \right]_{u=1}^{U} \\
&- \frac{1}{T} \left[\sum_{j=1}^{U} \mathbb{E} \left[\int_0^T K_{j,u} \lambda_u(t) h\left(\lambda_j(t)\right) dt \right] \right]_{u=1}^{U}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{T}\left[\sum_{j=1}^{U} K_{j,u}\int_0^T g(T-t)(M_{j,u}(t)-\alpha_t^{j,u}+\alpha_t^{j,u})dt\right]_{u=1}^{U} \\
&\quad - \frac{1}{T}\left[\sum_{j=1}^{U}\mathbb{E}\left[\int_0^T K_{j,u}\lambda_u(t)h\left(\lambda_j(t)\right)dt\right]\right]_{u=1}^{U} \\
&= \frac{1}{T}\left[\sum_{j=1}^{U} K_{j,u}\int_0^T g(T-t)(M_{j,u}(t)-\alpha_t^{j,u})dt\right]_{u=1}^{U} \\
&\quad + \frac{1}{T}\left[\sum_{j=1}^{U} K_{j,u}\left(\int_0^T \alpha_t^{j,u}g(T-t)dt - \mathbb{E}\left[\int_0^T \lambda_j(t)h\left(\lambda_j(t)\right)dt\right]\right)\right]_{u=1}^{U}
\end{aligned}
\tag{12}
$$

Majorize the first term using the assumption that $K_{j,u}$ is bounded by some number, say C, for all j, u. Jensen's Inequality can be applied using the fact that g is probability density function. Finally, with the assumption on h, $\mathbb{E}\left[(M_t^u-\alpha_t^{j,u})^2\right] = \mathbb{E}\left[\langle M_\bullet^u, M_\bullet^u\rangle_t\right] = \alpha_t^{j,u}$. Using these details, there is

$$
\begin{aligned}
&\mathbb{E}\left[\left(\frac{1}{T}\left[\sum_{j=1}^{U} K_{j,u}\int_0^T g(T-t)(M_{j,u}(t)-\alpha_t^{j,u})dt\right]_{u=1}^{U}\right)^2\right] \\
&\leq \frac{C^2}{T^2}\left[\mathbb{E}\left[\left(\sum_{j=1}^{U}\int_0^T g(T-t)(M_{j,u}(t)-\alpha_t^{j,u})dt\right)^2\right]\right]_{u=1}^{U} \\
&= \frac{C^2}{T^2}\left[\sum_{j=1}^{U}\mathbb{E}\left[\left(\int_0^T (M_{j,u}(t)-\alpha_t^{j,u})g(T-t)dt\right)^2\right]\right]_{u=1}^{U} \\
&\leq \frac{C^2}{T^2}\left[\sum_{j=1}^{U}\int_0^T \mathbb{E}\left[(M_{j,u}(t)-\alpha_t^{j,u})^2\right]g(T-t)dt\right]_{u=1}^{U} \\
&\leq \frac{C^2}{T^2}\left[\sum_{j=1}^{U}\int_0^T \alpha_t^{j,u}g(T-t)dt\right]_{u=1}^{U}
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&T\mathbb{E}\left[\left(\frac{1}{T}\left[\sum_{j=1}^{U} K_{j,u}\int_0^T g(T-t)(M_{j,u}(t)-\alpha_t^{j,u})dt\right]_{u=1}^{U}\right)^2\right] \\
&\leq \frac{C^2}{T}\left[\sum_{j=1}^{U}\int_0^T g(T-t)\alpha_t^{j,u}dt\right]_{u=1}^{U}.
\end{aligned}
\tag{13}
$$

According to the assumption 2, $\int_0^T g(T-t)\alpha_t^{j,u}dt$ is finite between any two sub-processes. Thus, (13) is also finite. It is shown that every other term is indeed finite as such, to find bounds for (10), the query boils down to finding the whole square bound for the second term in equation (12). To that extent, we use the upper-bound for $K_{j,u}$ and then apply integration by parts. Namely,

$$
\begin{aligned}
&T\mathbb{E}\left[\left(\frac{1}{T}\left[\sum_{j=1}^{U} K_{j,u}\left(\int_0^T \alpha_t^{j,u} g(T-t)dt - \mathbb{E}\left[\int_0^T \lambda_u(t)h\left(\lambda_j(t)\right)dt\right]\right)\right]_{u=1}^{U}\right)^2\right] \\
&\leq TC^2\left[\sum_{j=1}^{U}\mathbb{E}\left[\left(\frac{\alpha_T^{j,u}}{T}\int_0^T g(T-t)dt - \frac{1}{T}\int_0^T \frac{\partial\alpha_t^{j,u}}{\partial t}g(T-t)dt\right.\right.\right. \\
&\qquad\left.\left.\left. - \frac{1}{T}\mathbb{E}\left[\int_0^T \lambda_u(t)h\left(\lambda_j(t)\right)dt\right]\right)^2\right]\right]_{u=1}^{U}
\end{aligned}
\tag{14}
$$

The second term is always negative. This allows us to obtain a tighter bound for the expression. However, it cannot be proceeded on without specific guesses for h. □

Remark 2 In the standard univariate recursive case [28], the compensator $\alpha_t^{j,u} = t$ and the recursive function $h(x) = \frac{1}{x}$. In this particular case, the above expression (14) decomposes to

$$
\begin{aligned}
&TC^2\left[\mathbb{E}\left[\left(\int_0^T g(t)dt - \frac{1}{T}\int_0^T g(t)dt - 1\right)^2\right]\right] \\
=&TC^2\mathbb{E}\left[\left(\int_T^\infty g(t)dt + \frac{1}{T}\int_0^T g(t)dt\right)^2\right] \\
\leq&2C^2\left(\left(\sqrt{T}\int_T^\infty g(t)dt\right)^2 + \frac{1}{T}\left(\int_0^T g(t)dt\right)^2\right).
\end{aligned}
$$

With the assumption that $\limsup_{T\to\infty} \sqrt{T} \int_T^\infty g(t)dt < \infty$, it implies a uniform upper bound in t for the second term. Combining this bound with the results of (13), (11), the variance is finite and obtain an LLN result.

While a uniform bound is not arrived, a couple of "desired" properties of the recursive function h have been obtained in this section.

3 Parameter Fitting and Simulation Algorithms

3.1 Parameter Fitting Algorithms

The spatiotemporal multivariate recursive Hawkes processes (MRHPs) ($u = 1, 2, \cdots, U$) have the conditional event intensity function λ_u as the following:

$$\lambda_u(x, y, t) = \mu_u(x, y) + \sum_{t_i<t} H(\lambda_{u_i}(x_i, y_i, t_i), u_i, u) g(x-x_i, y-y_i, t-t_i), \quad (15)$$

where t_i and (x_i, y_i) are the timestamps and geo-locations, respectively; μ_u is the background rate for each state u; H is the effective reproduction number function, which depends on the event intensity at a given time; and g is the distribution function, called triggering kernel in Hawkes model.

Raise two assumptions:

1. $H(\lambda, u, v) = K_{u,v}h(\lambda)$.
2. $g(x, y, t) = g_1(t)g_2(x, y)$, where $g_1(t)$ is a one-dimensional distribution function about temporal value t, and $g_2(x, y)$ is a two-dimensional distribution about spatial value (x, y).

For a spatiotemporal model, the set of observations is a sequence $X = \{t_i, x_i, y_i, u_i\}_{i=1}^N$ with N events, where t_i is the timestamp when the given event happens. The spatial location of event i is then denoted as (x_i, y_i), and $u_i \in \{1, 2, \ldots, U\}$ is the category index of event i. The goal of parameter fitting is to find the parameters and functions, such as $K = \{K_{u,v}\}_{u,v=1}^U$, $\mu = \{\mu_u(x, y)\}_{u=1}^U$, $g_1(t)$, $g_2(x, y)$, and $h(\lambda)$ that fit to the observation sequence $\{t_i, x_i, y_i, u_i\}_{i=1}^N$ the best. In other words, it needs to maximize the log-likelihood function over these parameters:

$$\{\mu_u\}_{u=1}^U, \{K_{u,v}\}_{u,v=1}^U, h(\cdot), g_1(\cdot), g_2(\cdot,\cdot) = \underset{\mu,K,h,g_1,g_2}{\arg\max}\ L(\mu, K, h, g_1, g_2; X), \quad (16)$$

where $L(\mu, K, h, g_1, g_2)$ is the log-likelihood function that

$$L(\mu, K, h, g_1, g_2; X) = \sum_{i=1}^{N} \log(\lambda_{u_i}(x_i, y_i, t_i)) - \sum_{u=1}^{U} \int_0^T \int_{\mathbb{R}^2} \lambda_u(x_i, y_i, t_i) dx dy dt. \tag{17}$$

Using information about the functional form of the recursive function $h(\lambda)$ and the triggering functions $g_1(t), g_2(x, y)$, one can fit the model parametrically. If, however, these functional forms are unknown, the algorithm can be extended to semi-parametric or non-parametric versions as in Yuan et al. [33].

3.1.1 Parametric (or Semi-parametric) Estimation

In this section, it is assumed that the form of triggering functions g_1, g_2 is known, and the model can be fitted no matter whether the information of recursive function h is given or not. As an example, consider a multivariate recursive Hawkes process with a specific parametric form:

- $\mu_u(x, y)$ is a linear combination of a series of two-dimensional Gaussian kernels i.i.d. $\sim \mathcal{N}(\mathbf{0}, \eta^2 I_2)$, where I_2 is the 2×2 identity matrix:

$$\mu_u(x, y) = \sum_{i=1}^{N} \frac{\beta_{u_i,u}}{2\pi\eta^2 T} \exp\left(-\frac{(x - x_i)^2 + (y - y_i)^2}{2\eta^2}\right). \tag{18}$$

- $g_1(t)$ is an exponential decaying function with mean $\frac{1}{\omega}$:

$$g_1(t) = \omega e^{-\omega t}. \tag{19}$$

- $g_2(x, y)$ is a two-dimensional Gaussian distribution $\mathcal{N}(\mathbf{0}, \sigma^2 I_2)$:

$$g_2(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right). \tag{20}$$

- If $h(\lambda)$ is fitted parametrically, it is assumed to be an exponential decaying function:

$$h(\lambda) = \alpha e^{-\alpha\lambda}. \tag{21}$$

Otherwise h is fitted non-parametrically as a piecewise constant function:

$$h(\lambda) = h_i, \ \forall \lambda \in D_i, \tag{22}$$

where $\{D_i\}_{i=1}^{N_\lambda}$ is a set of bins in $\mathbb{R}$.

Here, T denotes the time span of a data set. The quantity $\beta_{u_i,u}$ denotes the extent to which events in process u_i contribute to the background rate for events in process u.

The EM algorithm [30] is applied to estimate those parameters since EM type of algorithms are popular approaches to maximum-likelihood optimization problem with incomplete data.

In the presented point process model, it only provides the observations about the category and spatiotemporal information of each event with any information about the branching structure of a point process. In this sense, the data set is incomplete.

It comes to the estimation of parameter Θ from data X. One can view the likelihood function $L(\Theta; X)$ as the marginal likelihood function of $L(\Theta; Y, X)$, where Y is a latent variable. $L(\Theta; Y, X)$ is referred as the "complete-datalikelihood function," while $L(\Theta; X)$ is the "incomplete-data likelihood function." Since both Y and $L(\Theta; Y, X)$ are random variables, they cannot be estimated directly. Therefore, consider the following expectation function:

$$\begin{aligned} Q(\Theta, \Theta^{i-1}) &= \mathbb{E}[\log(L(\Theta; Y, X)|X, \Theta^{i-1}] \\ &= \int \log(L(\Theta; Y, X) f(Y|X, \Theta^{i-1}) dY, \end{aligned} \tag{23}$$

where $f(Y|X, \Theta^{i-1})$ is the conditional probability density function of Y, given the data X and Θ^{i-1}. Then update parameters by solving the following equation:

$$\hat{\Theta}^i = \arg\max_{\Theta} Q(\Theta, \Theta^{i-1}). \tag{24}$$

In practice, given the spatial region R and the time window $[0, T]$, the log-likelihood function is

$$L(\Theta; X) = \sum_{i=1}^{N} \log(\lambda_{u_i}(x_i, y_i, t_i)) - \sum_{u=1}^{U} \int_0^T \int_R \lambda_u(x_i, y_i, t_i) dx dy dt, \tag{25}$$

where X is the observation sequence $\{(t_i, x_i, y_i, u_i)\}_{i=1}^N$ and Θ is the parameters μ, K, g_1, g_2, h.

Consider latent variables $Y_{i,j}$ and $Y_{i,j}^b$ as

$$Y_{i,j} = \begin{cases} \mathbb{1}_{\{\text{Event i is triggered by event j via triggering kernel } g\}}, & i > j \\ 0, & i \leq j \end{cases}$$

$$Y_{i,j}^b = \mathbb{1}_{\{\text{Event i is triggered by event j via background rate } \mu\}}, \quad i \geq j$$

where $\mathbb{1}_P$ is an indicator function that

$$\mathbb{1}_P = \begin{cases} 1, & \text{if P is true} \\ 0, & \text{if P is false.} \end{cases} \tag{26}$$

Then, define two probability matrices $P = \{P_{i,j}\}_{i,j=1}^{N} = \{\mathbb{E}[Y_{i,j}]\}_{i,j=1}^{N}$ and $P^b = \{P_{i,j}^b\}_{i,j=1}^{N} = \{\mathbb{E}[Y_{i,j}^b]\}_{i,j=1}^{N}$. The complete log-likelihood function is as the following:

$$\begin{aligned} L(\Theta; Y, Y^b, X) = & \sum_{i=1}^{N}\sum_{j=1}^{N} Y_{i,j}^b \log(\mu_{u_i}) - \sum_{u=1}^{U}\sum_{i=1}^{N} \beta_{u_i,u} \\ & + \sum_{i=2}^{N}\sum_{j=1}^{i-1} Y_{i,j} \log[K_{u_j,u_i} g(x_i - x_j, y_i - y_j, t_i - t_j) h(\lambda_{u_j}(x_j, y_j, t_j))] \\ & - \sum_{u=1}^{U} \int_0^T \int_{\mathbb{R}^2} \sum_{t_i < t} K_{u_i,u} h(\lambda_{u_i}(x_i, y_i, t_i)) g(x - x_i, y - y_i, t - t_i) dx dy dt. \end{aligned} \tag{27}$$

The expectation function is formulated by:

$$\begin{aligned} Q(\Theta) = & \sum_{i=1}^{N}\sum_{j=1}^{N} P_{i,j}^b \log\left(\frac{\beta_{u_i,u}}{2\pi\eta^2 T} \exp(-\frac{(x - x_i)^2 + (y - y_i)^2}{2\eta^2})\right) - \sum_{u=1}^{U}\sum_{i=1}^{N} \beta_{u_i,u} \\ & + \sum_{i=2}^{N}\sum_{j=1}^{i-1} P_{i,j} \log\left[K_{u_j,u_i} h(\lambda_{u_j}(x_j, y_j, t_j))\right] \\ & + \sum_{i=2}^{N}\sum_{j=1}^{i-1} P_{i,j} \log\left[\frac{\omega}{2\pi\sigma^2} \exp\left(-\omega(t_i - t_j) - \frac{(x_i - x_j)^2 + (y_i - y_j)^2}{2\sigma^2}\right)\right] \\ & - \sum_{u=1}^{U}\sum_{i=1}^{N} K_{u_i,u} h\left(\lambda_{u_i}(x_i, y_i, t_i)\right)\left(1 - e^{-\omega(T - t_i)}\right) \end{aligned} \tag{28}$$

In the maximization step of EM algorithm, it needs to solve the equation:

$$\hat{\Theta} = \arg\max_{\Theta} Q(\Theta). \tag{29}$$

In the model $\Theta = \{\mu, K, h, g_1, g_2\} = \{\beta, \eta, K, h, \omega, \sigma\}$, where $\beta = \{\beta_{u,v}\}_{u,v=1}^{U}$, $K = \{K_{u,v}\}_{u,v=1}^{N}$ and $\eta, \alpha, \omega, \sigma$ are positive scalars. In practice, the maximization step is performed directly by taking derivatives with respect to the parameters and setting them to 0, and parameters are optimized one by one while fixing other parameters.

As for the expectation step, since $Y_{i,j}$ and $Y_{i,j}^b$ denote the triggering relationship from triggering kernel g and the background rate μ, respectively, they can be

estimated from the previous estimation of $\Theta = \{\mu, K, g_1, g_2, h\}$ according to the conditional intensity formula (15) as

$$P_{i,j} = \begin{cases} K_{u_j,u_i} h(\lambda_{u_j}(x_j, y_j, t_j)) g_1(t_i - t_j) g_2(x_i - x_j, y_i - y_j)/\lambda(x_i, y_i, t_i), & i > j \\ 0, & i \leq j. \end{cases}$$

$$P^b_{i,j} = \beta_{u_j,u_i} \exp\left(-\frac{(x_j - x_i)^2 + (y_j - y_i)^2)}{2\eta^2}\right) / (2\pi\eta^2 T \lambda(x_i, y_i, t_i)).$$

By iterating the expectation and maximization steps alternatively, the overall Algorithm 1 is developed for the parametric model. Note that as for the initialization, sample $\beta = \{\beta_{u,v}\}_{u,v=1}^{U}$, $K = \{K_{u,v}\}_{u,v=1}^{N}$ and h, g_1, g_2 at random according to their parametric (non-parametric) forms.

Algorithm 1 is the EM-based algorithm to parametrically (semi-parametrically) approximate parameters in the MRHP model.

Remark 3 (Update the Recursive Function h) If $h(\lambda) = \alpha e^{-\alpha\lambda}$, in order to simplify the calculation, practically consider $\tilde{h}(\lambda) = e^{-\alpha\lambda}$. By taking the derivative of the expectation function (28), α can be updated by solving the following equation:

$$\sum_{u=1}^{U}\sum_{i=1}^{N} K_{u_i,u} \lambda_{u_i}(x_i, y_i, t_i) \exp\left(-\alpha\lambda_{u_i}(x_i, y_i, t_i)\right) [1 - \exp(-w(T - t_i))] - \sum_{j<i} P_{i,j} \lambda_{u_j}(x_j, y_j, t_j) = 0 \tag{30}$$

In practice, the solution α of (30) can be approximated by the "fzero" function in MATLAB, which is a combination of bisection, secant, and inverse quadratic interpolation methods. The old α is used to be the initial guess of this iterative scheme. If the output α is a negative number, set $\alpha = 0$ instead. With the output α and K, set $h(\lambda) = \alpha e^{-\alpha\lambda}$ and $K = K/\alpha$ as the final result.

h can be fitted non-parametrically without any information about it. Let $m = \min_{1\leq i\leq N} \lambda_i$ and $M = \min_{1\leq i\leq N} \lambda_i$. Set up N_λ bins $D_k = [d_{k-1}, d_k)$, $k = 1, 2, \ldots, N_\lambda$ with $d_k = m + k((M - m)/N_\lambda$. h is assumed to be a piecewise constant function. In each bin D_k, h has a constant value h_k defined by:

$$h_k = \begin{cases} \frac{\sum_{i=2}^{N}\sum_{j=1}^{i-1} P_{i,j} \mathbb{1}_{\{\lambda_{u_j}(x_j, y_j, t_j) \in D_k\}}}{\sum_{u=1}^{U}\sum_{i=1}^{N} K_{u,u_i} \mathbb{1}_{\{\lambda_{u_i}(x_i, y_i, t_i) \in D_k\}} [1 - \exp(\omega(T - t_i))]}, & \{\lambda_{u_j}(x_j, y_j, t_j)\}_{i=1}^{N} \cap D_k \neq \emptyset, \\ 0, & \text{otherwise} \end{cases} \tag{31}$$

Finally, normalize h such that

$$\sum_{k=1}^{N_\lambda} |D_k| h_k = 1. \tag{32}$$

Algorithm 1: Parametric (Semi-parametric) solver for the spatiotemporal multivariate recursive Hawkes model

Inputs: Point process $\{(t_i, x_i, y_i, u_i)\}_{i=1}^N$ with time window $[0, T]$. Termination threshold: ϵ. Maximum number of iteration: $I_{\max}$

Initialization: Randomly sample two $U \times U$ matrices as $K^{(0)}$ and $\beta^{(0)}$ according to $\mathcal{U}[0, 1]$. Randomly sample 2 scalars as $\omega^{(0)}, \sigma^{(0)}$ according to $\mathcal{U}[0, 1]$ and let $\eta^{(0)} = \sigma^{(0)}$. Choose $h^{(0)}(\lambda) = e^{-\alpha^{(0)}\lambda}$ with $\alpha^{(0)} \sim \mathcal{U}[0, 1]$ if h is parametric, otherwise choose $h^{(0)}(\lambda) \equiv 1$ (constant function).

Outputs: Model parameters $\Theta = \{\beta, K, \eta, \omega, \sigma, h\}$.

Initialize $\delta = 1$ and $l = 0$.

while $\delta > \epsilon$ *and* $l < I_{max}$ **do**

Updates

1. Update background rates $\mu_{u_i}^{(l)}(x_i, y_i),\ i = 1, 2, \ldots, N$:

$$\mu_{u_i}^{(l)}(x_i, y_i) = \sum_{j=1}^{N} \frac{\beta_{u_j,u_i}^{(l)}}{2\pi\eta^{2,(l)}T} \exp\left(-\frac{(x_i - x_j)^2 + (y_i - y_j)^2}{2\eta^{2,(l)}}\right).$$

2. Update triggering functions:

$$g_1^{(l)}(t) = \omega^{(l)} e^{-\omega^{(l)} t},$$
$$g_2^{(l)}(x, y) = \frac{1}{2\pi(\sigma^{(l)})^2} \exp\left(-\frac{x^2 + y^2}{2(\sigma^{(l)})^2}\right).$$

3. Update intensity $\lambda_{u_i}^{(l)}(x_i, y_i, t_i), i = 1, 2, \ldots, N$:

$$\lambda_{u_i}^{(l)}(x_i, y_i, t_i) = \mu_{u_i}^{(l)}(x_i, y_i) + \sum_{t_j < t_i} K_{u_j,u_i}^{(l)} h^{(l)}(\lambda_{u_j}^{(l)}(x_j, y_j, t_j)) g_1^{(l)}(t_i - t_j) g_2^{(l)}(x_i - x_j, y_i - y_j).$$

Expectation step:

1. Estimate triggering probabilities, for $i, j = 1, 2, \ldots, N$:

$$P_{i,j}^{(l)} = \begin{cases} K_{u_j,u_i}^{(l)} h^{(l)}(\lambda_{u_j}^{(l)}(x_j, y_j, t_j)) g_1^{(l)}(t_i - t_j) g_2^{(l)}(x_i - x_j, y_i - y_j), & i > j \\ 0, & i \le j. \end{cases}$$

$$P_{i,j}^{b,(l)} = \beta_{u_j,u_i}^{(l)} \exp\left(-\frac{(x_j - x_i)^2 + (y_j - y_i)^2)}{2\eta^{2,(l)}}\right) / (2\pi\eta^{2,(l)}T).$$

2. Normalize $P_{i,j}^{(l)}, P_{i,j}^{b,(l)}$ such that $\sum_{j=1}^{N} \left[P_{i,j}^{(l)} + P_{i,j}^{b,(l)}\right] = 1,\ \forall i = 1, 2, \cdots, N.$

Maximization step:

1. $\omega^{(l+1)} = \frac{\sum_{j<i} P_{i,j}^{(l)}}{\sum_{j<i} P_{i,j}^{(l)}(t_i - t_j) + \sum_{u=1}^{U} \sum_{i=1}^{N} K_{u_j,u}^{(l)} h^{(l)}[\lambda_{u_j}^{(l)}(x_i, y_i, t_i)][(T - t_i)\exp(-\omega^{(l)}(T - t_i))]}$.

2. Update $h^{(l+1)}$ (details in Remark 3).

3. $K_{u,v}^{(l+1)} = \frac{\sum_{u_i=v} \sum_{u_j=u} P_{i,j}^{(l)}}{\sum_{u_j=u} h^{(l+1)}[\lambda_{u_j}^{(l)}(x_j, y_j, t_j)][1 - \exp(-w^{(l+1)}(T - t_j))]},\ u, v = 1, 2, \ldots, U.$

4. $\beta_{u,v}^{(l+1)} = \frac{\sum_{u_i=v} \sum_{u_j=u} P_{i,j}^{b,(l)}}{\sum_{u_j=u} 1},\ u, v = 1, 2, \ldots, U.$

5. $(\sigma^{(l+1)})^2 = \frac{\sum_{i,j=1}^{N} (P_{i,j}^{(l)} + P_{i,j}^{b(l)})((x_i - x_j)^2 + (y_i - y_j)^2)}{2\sum_{i,j=1}^{N} (P_{i,j}^{(l)} + P_{i,j}^{b,(l)})}$, $(\eta^{(l+1)})^2 = (\sigma^{(l+1)})^2$.

$l = l + 1$,
$\delta = \max\{\|K^{(l+1)} - K^{(l)}\|_\infty, \|\beta^{(l+1)} - \beta^{(l)}\|_\infty, |\omega^{(l+1)} - \omega^{(l)}|, |\sigma^{(l+1)} - \sigma^{(l)}|\}$.

end

Remark 4 (Guaranty of Positive) According to the model assumptions, parameters $\mu, \omega, \sigma, P_{i,j}, P^b_{i,j}, K_{u,v}, \beta_{u,v}$ should all be non-negative numbers. In fact, according to the method of random initialization and iteration steps in Algorithm 1, these parameters are guaranteed to be non-negative.

3.1.2 Temporal Version of Parameter Fitting Algorithms

In Sect. 5, the temporal MRHP model is implemented to deal with real daily reported cases data of COVID-19. The corresponding temporal version of parameter fitting algorithm is slightly different from the spatiotemporal case. In the temporal model, the conditional intensity of sub-process u degenerates to:

$$\lambda_u(t) = \mu_u + \sum_{t_i<t} H(\lambda_{u_i}(t_i), u_i, u) g(t - t_i) = \mu_u + \sum_{t_i<t} K_{u_i,u} h(\lambda_{u_i}(t_i)) g(t - t_i). \tag{33}$$

Let $X = \{(t_i, u_i)\}_{i=1}^N$ be the observed point process. The model settings are the same as those in Sect. 3.1.1. Considering the latent variables

$$Y_{i,j} = \begin{cases} \mathbb{1}_{\{\text{Event i is triggered by event j}\}}, & i > j \\ \mathbb{1}_{\{\text{Event i is a background event}\}}, & i = j \\ 0, & i < j, \end{cases} \tag{34}$$

and the probability matrix $P = \{P_{i,j}\}_{i,j=1}^N = \{\mathbb{E}(Y_{i,j})\}_{i,j=1}^N$, the corresponding complete log-likelihood function can be derived as

$$\begin{aligned} L(\mu, K, h, g; Y, X) &= \sum_{i=1}^N Y_{i,i} \log(\mu_{u_i}) + \sum_{i=2}^N \sum_{j=1}^{i-1} Y_{i,j} \log[K_{u_j,u_i} g(t_i - t_j) h(\lambda_{u_j}(t_j))] \\ &\quad - \sum_{u=1}^U \int_0^T [\mu_u + \sum_{t_i<t} K_{u_i,u} g(t - t_i) h(\lambda_{u_i}(t_i))] dt \\ &= \sum_{u=1}^U (\sum_{u_i=u} Y_{i,i}) \log(\mu_u) - T \sum_{u=1}^U \mu_u \\ &\quad + \sum_{i=2}^N \sum_{j=1}^{i-1} Y_{i,j} \log[K_{u_j,u_i} g(t_i - t_j) h(\lambda_{u_j}(t_j))] \\ &\quad - \sum_{u=1}^U \int_0^T \sum_{t_i<t} K_{u_i,u} g(t - t_i) h(\lambda_{u_i}(t_i)) dt. \end{aligned} \tag{35}$$

The temporal version of parameter fitting algorithm 2 for both parametric and semi-parametric models can be built similarly as the process in Sect. 3.1.1.

Algorithm 2: Parametric (Semi-parametric) solver for the temporal multivariate recursive Hawkes model

Inputs: Point process $\{(t_i, u_i)\}_{i=1}^{N}$. Termination threshold: ϵ. Maximum number of iteration: $I_{\max}$.

Initialization: Randomly sample a $U \times U$ matrices as $K^{(0)}$ according to $\mathcal{U}[0, 1]$. Randomly sample $U + 1$ scalars as $\mu^{(0)} = \{\mu_u^{(0)}\}_{u=1}^{U}$ and $\omega^{(0)}$ according to $\mathcal{U}[0, 1]$. Choose $h^{(0)}(\lambda) = e^{-\alpha^{(0)}\lambda}$ with $\alpha^{(0)} \sim \mathcal{U}[0, 1]$ if h is parametric, otherwise choose $h^{(0)}(\lambda) \equiv 1$ (constant function).

Outputs: Model parameters $\Theta = \{\mu, K, h, g\}$.

Initialize $\delta = 1$ and $l = 0$.

while $\delta > \epsilon$ **do**

Updates

1. Update the triggering function: $g^{(l)}(t) = \omega^{(l)} e^{-\omega^{(l)} t}$.
2. Update condition intensity $\lambda_{u_i}^{(l)}(t_i),\ i = 1, 2, \ldots, N$ by

$$\lambda_u^{(l)}(t_i) = \mu_u^{(l)} + \sum_{t_j < t_i} K_{u_j,u}^{(l)} h^{(l)}(\lambda_{u_j}^{(l)}(t_j)) g^{(l)}(t_i - t_j).$$

Expectation step:

1. $P_{i,j}^{(l)} = \begin{cases} K_{u_j,u_i}^{(l)} h^{(l)}(\lambda_{u_j}(t_j)) g^{(l)}(t_i - t_j),\ i > j \\ \mu_{u_i}^{(l)},\ i = j \\ 0,\ i < j. \end{cases}$.
2. Normalize $P_{i,j}^{(l)}$ such that $\sum_{j=1}^{N} P_{i,j}^{(l)} = 1,\ \forall i = 1, 2, \cdots, N$.

Maximization step:

1. $\mu_u^{(l+1)} = \sum_{u_i = u} P_{i,i}/T\ \ u = 1, 2, \ldots, U$.
2. $\omega^{(l+1)} = \frac{\sum_{j<i} P_{i,j}^{(l)}}{\sum_{j<i} P_{i,j}^{(l)}(t_i - t_j) + \sum_{u=1}^{U} \sum_{i=1}^{N} K_{u_i,u}^{(l)} h^{(l)}[\lambda_{u_i}^{(l)}(t_i)][(T - t_i)\exp(-\omega^{(l)}(T - t_i))]}$.
3. Update $h^{(l+1)}$ according to Remark 3.
4. $K_{u,v}^{(l+1)} = \frac{\sum_{u_i = v} \sum_{u_j = u} P_{i,j}^{(l)}}{\sum_{u_j = u} h^{(l+1)}[\lambda_{u_j}^{(l)}(t_j)][1 - \exp(-w^{(l+1)}(T - t_j))]},\ u, v = 1, 2, \ldots, U$.

$l = l + 1, \delta = \max\{\|K^{(l+1)} - K^{(l)}\|_\infty, |\omega^{(l+1)} - \omega^{(l)}|\}$.

end

3.2 Simulation Algorithm

To generate synthetic data for model comparisons and verification of parameter fitting algorithm, it needs to simulate the multivariate recursive Hawkes point processes with the conditional intensity in (15) for each sub-process u. Branching structure method [33, 35] of self-exciting point processes is applied to develop Algorithm 3 for simulations.

Algorithm 3: Simulation of a spatiotemporal multivariate recursive Hawkes process

Inputs: Time window size T and spatial region R. Background rate $\mu = \{\mu_u\}_{u=1}^{U}$. Temporal and spatial triggering kernels (distribution function) $g_1(t)$, $g_2(x, y)$. Recursive function $h(\lambda)$.

Outputs: Point process $H = \{(t_i, x_i, y_i, u_i, \lambda_i)\}_{i=1}^{N}$, where (t_i, x_i, y_i) are the spatiotemporal stamps of event i, u_i is the category index of event i and $\lambda_i = \lambda_{u_i}(x_i, y_i, t_i)$ is the conditional intensity of event i.

Initialization: An empty set H and an empty stack Q.

Generate background events:

1. Draw N_u^b, the number of background events of type u, from a Poisson distribution with parameter $\theta = \mu_u T$ for each $u \leq U$.
2. Add each background event $i \leq \sum_{u=1}^{U} N_u^b$, with its associated tuple (t_i, x_i, y_i, u_i), to the stack Q, with t_i and (x_i, y_i) drawing from the uniform spatiotemporal distribution over the time interval $[0, T]$ and a bounded spatial region R respectively.

Generate triggered events:
while Q *is not empty* **do**

1. Pick the event i: (t_i, x_i, y_i, u_i) in stack Q with the minimal time and remove this event from stack Q.
2. Calculate the conditional intensity related to event i by

$$\lambda_{u_i}(x_i, y_i, t_i) = \mu_u(x_i, y_i) + \sum_{j \in H} K_{u_j, u_i} h(\lambda_j) g_1(t_i - t_j) g_2(x_i - x_j, y_i - y_j).$$

3. Add $(t_i, x_i, y_i, u_i, \lambda_i)$ into set H.
4. Draw n_i, the number of events triggered by event i, from a Poisson distribution with parameter $\theta_i = \sum_{u=1}^{U} K_{u_i, u} h(\lambda_i)$.
5. Generate events (t_k, x_k, y_k, u_k) for each $k \leq n_i$, where t_k, (x_k, y_k) and u_k are sampled from $g_1(t)$, $g_2(x, y)$ and $P(u_k = \tilde{u}) = \frac{K_{u_i, \tilde{u}}}{\sum_{u=1}^{U} K_{u_i, u}}$.
6. Add (t_k, x_k, y_k, u_k) to stack Q if $t_k < T$.

end

Remark 5 (Temporal Version of Simulation Algorithm) If it only needs to simulate a temporal point process, just replace the spatiotemporal intensity formula (15) by the temporal intensity (33) and drop everything else about the spatial coordinates (x_i, y_i) in Algorithm 3.

4 Reconstruct Multivariate Point Process from Data with Imprecise Time

In practice, it may be difficult to obtain data with very precise happening time of event times. For example, in COVID-19 data, only the number of infections happening on each date is provided without the precise infection time of each event.

According to experiments in Sect. 5, the fitted parameters can be really far away from the ground truth if the model from dataset is fitted with imprecise time.

In this section, the original point process is approximately reconstructed from a point process with imprecise time. Since there is only the temporal bias, consider a temporal point process $\widetilde{H} = \{(\tilde{t}_i, \tilde{u}_i)\}_{i=1}^{N}$ as the observation data with imprecise time. The goal is to reconstruct its original point process $H = \{t_i, u_i\}_{i=1}^{N}$.

The imprecise time steps are assumed to be derived from rounding up the result of the original time t_i due to reports being issued daily. Also, the testing results may have some delay in different states. Therefore, purely inferred the parameters from the COVID data can lead to the wrong conclusion.

4.1 Time Reconstruction

The shape-preserving piecewise cubic interpolation method is applied for the time reconstruction. It can be implemented by Akima's algorithm [1, 2]. This algorithm performs cubic interpolation to produce piecewise polynomials with continuous first-order derivatives (C^1) and preserves the slope and avoids undulations in flat regions. A flat region occurs whenever there are three or more consecutive collinear points, which are connected by the algorithm with a straight line. To ensure that the region between two data points is flat, insert an additional data point between those two points.

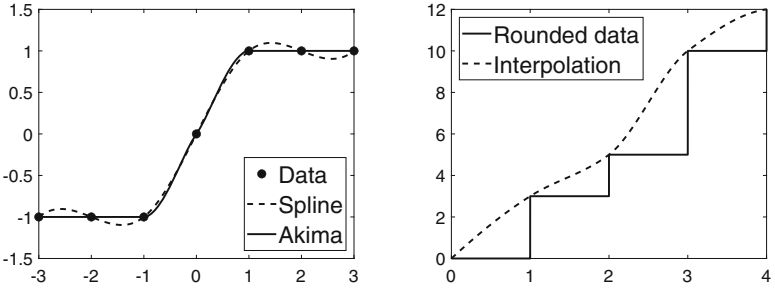
When two flat regions with different slopes meet, the modification made by the original Akima algorithm gives more weight to the side where the slope is closer to zero. This modification gives priority to the side that is closer to horizontal, which is more intuitive and avoids the overshoot, and the original Akima algorithm gives equal weights to the points on both sides, thus evenly dividing the undulation.

The spline algorithm, on the other hand, performs cubic interpolation to produce piecewise polynomials with continuous second-order derivatives (C^2). The result is comparable to a regular polynomial interpolation but is less susceptible to heavy oscillation between data points for high degrees. Compared to the spline algorithm, the Akima algorithm produces fewer undulations and is better suited to deal with quick changes between flat regions. This difference is illustrated in Fig. 1a.

Such an interpolation method is chosen to reconstruct the cumulative curve because of two reasons:

- The input cumulative function of imprecise time data is an increasing function. The interpolation result is supposed to preserve the increasing property. Hence, a shape-preserving method is needed.
- Recall the definition of conditional intensity λ of a given counting process $N(t)$:

$$\lambda(t) = \lim_{t \to 0} \frac{P\{N(t+h) - N(t) = 1 | \mathcal{H}_t\}}{h},$$



(a) (b)

Fig. 1 Panel (**a**): Difference between the shape-preserving interpolation (Akima) and the cubic spline interpolation. It can be seen that the shape-preserving interpolation is more suitable for the present problem. Panel (**b**): Reconstruct the curve base on a step function through the shape-preserving interpolation

where $\mathcal{H}_t$ is the history of point process before time t. This definition is similar to the first derivative. It implies to require the interpolation result to be a C^1 function.

The time sequence reconstruction algorithm 4 is based on this shape-preserving piece cubic interpolation method.

Algorithm 4: Reconstruction of time sequence

Inputs: An integer-time sequence $\{\widetilde{T}_i\}_{i=1}^N$.

Outputs: A reconstructed time sequence $\{t_i\}_{i=1}^N$.

1. Calculate cumulative step function $F(t) = \sum_{i=1}^N \mathbb{1}_{\{t_i < t\}}$, where $\mathbb{1}$ is the indicator function.
2. Apply Akima's algorithm to get the shape-preserving piecewise cubic interpolation result based on nodes $\{n, F(n)\},\ n \le \max_i\{\widetilde{T}_i\},\ n \in \mathbb{N}$. (Illustrated by Fig. 1b)
3. Find the time steps $\{T_i\}_{i=1}^N$ when the interpolation curve reaches every integer. Here $T_i = \inf(F^{-1}([i, \infty))),\ i = 1, 2, \ldots, N$ (Illustrated by Fig. 2a)
4. Define $T_0 = 0$. For $i = 1, 2, \ldots, N$, if $T_i - T_{i-1} \le 1$, choose t_i as a uniformly random variable in $[T_{i-1}, T_i]$, otherwise choose t_i as a uniformly random variable in $[\lceil T_{i-1} \rceil - 1, T_i]$. The reconstructed time sequence is $\{t_i\}_{i=1}^N$. (Illustrated by Fig. 2b)

4.2 Category Index Reconstruction

In the multivariate point process, after the reconstruction of time sequence, it needs to find the category that each event in the reconstructed time sequence belongs to. The category index u_i is really important since the potential causal relationship

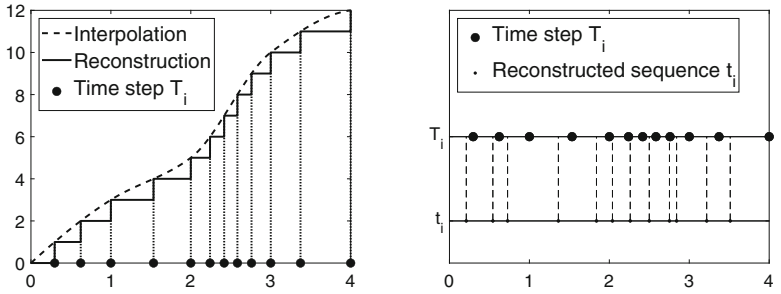


Fig. 2 Panel (**a**): Find the time steps $\{T_i\}_{i=1}^N$ from the interpolation curve. Panel (**b**): Reconstruct time sequence $\{t_i\}_{i=1}^N$ from time steps $\{T_i\}_{i=1}^N$. Details are in Algorithm 4

between two events is decided from their temporal values. If category indices are mixed, it might lead to inverse causal relationships or chaotic branching structure.

Define $\widetilde{H}_u = \{(\widetilde{T}_i, \tilde{u}_i)\}_{\tilde{u}_i=u}$, $u = 1, 2, \ldots, U$. Apply Algorithm 4 on the time sequences of $\widetilde{H}, \widetilde{H}_1, \ldots, \widetilde{H}_U$ to get the reconstructed time sequences $T = (t_i)_{i=1}^N$, $T_1 = (t_i^1)_{i=1}^{N_1}, \ldots, T_U = (t_i^U)_{i=1}^{N_U}$, where N_u is the number of events in sub-process $\widetilde{H}_u$ and $\sum_{u=1}^U N_u = N$.

The goal is to find an index sequence $\{u_i\}_{i=1}^N$ solves the following optimization problem:

$$\begin{aligned} &\min_{\{u_i\}_{i=1}^N} \sum_{u=1}^{U} \sum_{u_i=u} \|\widetilde{T}_u - T_u\|^2, \\ &s.t. \quad \sum_{u_i=u} 1 = N_u, \end{aligned} \tag{36}$$

where $\widetilde{T}_u$ is the sequence form by all $\tilde{t}_i$ that $\tilde{u}_i = u$. According to rearrangement inequality [15], the problem (36) can be solved explicitly. That is the category index reconstruction algorithm 5.

Algorithm 5: Reconstruction of category indices

Inputs: A temporal point process $\widetilde{H} = \{(\tilde{t}_i, \tilde{u}_i)\}_{i=1}^N$ with $\tilde{t}_i \in \mathbb{N}$.

Outputs: A reconstructed index sequence $\{u_i\}_{i=1}^N$.

1. Apply Algorithm 4 on the time sequences of $\widetilde{H}, \widetilde{H}_1, \ldots, \widetilde{H}_U$ to get the reconstructed time sequences $T = (t_i)_{i=1}^N$, $T_1 = (t_i^1)_{i=1}^{N_1}, \ldots, T_U = (t_i^U)_{i=1}^{N_U}$. It is illustrated by Fig. 3a.
2. Consider the composite point process $\hat{H} = \cup_{u=1}^U \{(t_i^u, u)\}_{i=1}^{N_u}$. Sort the events of $\hat{H}$ according to the time array from small to large. The resulting array of category indices $\{u_i\}_{i=1}^N$ is the solution to (36). It is illustrated by Fig. 3b.

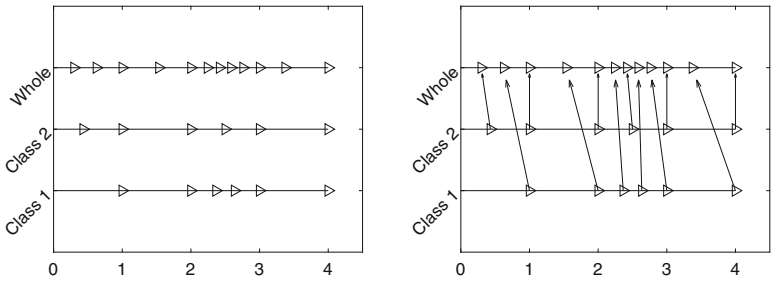


Fig. 3 Panel (**a**): Reconstruct time sequence about whole process and each sub-process. Panel (**b**): Reconstruct category indices by solving the minimal difference problem (36). The arrows indicate how to label the case in the whole process based on the reconstructed time sequence of sub-processes

Combining the reconstructed time sequence $\{t_i\}_{i=1}^N$ and reconstructed index sequence $\{u_i\}_{i=1}^N$, one can have the reconstructed point process $H = \{(t_i, u_i)\}_{i=1}^N$.

5 Numerical Experiments and Results

In this section, the parameter fitting algorithms are applied to both synthetic and real COVID-19 data sets. Experiments on synthetic data sets are used to verify the parameter fitting and simulation algorithms in Sect. 3. Meanwhile, the necessity of having a reconstructed algorithm mentioned in section is verified Sect. 4 via the synthetic data. Further experiments are made on real COVID-19 data. A MRHP model is fitted according to real data, and predictions are made based on the simulation algorithm 3. The predicted results are compared with the ground truth to test the performance of MRHP. The code to reproduce the experiments is available on Github.[1]

5.1 Synthetic Data Sets

In this part, under different choices of background rate μ, triggering function g, productivity matrix K, and recursive function h, a synthetic point process $X = \{(x_i, y_i, t_i, u_i)\}_{i=1}^N$ (if the point process is not spatiotemporal, it should be $X =$

[1] https://github.com/wispcarey/Bohan_MRHP.

$\{(t_i, u_i)\}_{i=1}^N$) is generated and applied to fit a MRHP model. The fitted parameters are compared with the ground truth under several metrics:

- For parametrically fitted triggering function g and recursive function h, the error metric is the **relative error** of the corresponding parameters.
- For background rate μ and productivity matrix K, the **relative error** of the 2-norm, i.e.,

$$\text{Error}(K) = \frac{\|K^{\text{fit}} - K\|_2}{\|k\|_2}, \ \text{Error}(\mu) = \frac{\|\mu^{\text{fit}} - \mu\|_2}{\|\mu\|_2}. \tag{37}$$

- In the recursive model, the productivity number $H(\lambda, u, v) = K_{u,v}h(\lambda)$. There might be the case that the matrix K and function h are fitted poorly but the fitting of $H(\lambda, u, v) = K_{u,v}h(\lambda)$ is relatively acceptable. Hence it is necessary to define a fitting error of the productivity number

$$\text{error}(H) = \frac{1}{N}\sum_{i=1}^{N} \frac{\|K^{\text{fit}}h^{\text{fit}}(\lambda_i) - Kh(\lambda_i)\|_2}{\|Kh(\lambda_i)\|_2}, \tag{38}$$

  where λ_i is the conditional intensity of the ith case of the point process X, i.e., $\lambda_i = \lambda(x_i, y_i, t_i)$ ($\lambda_i = \lambda(t_i)$ if the point process is not spatiotemporal).

For the temporal model, the parameter fitting algorithm terminates if the number of iterations reaches its maximum value, 1000, the difference of likelihood function (27) between a pair of neighbor steps is less than 10^{-6}, or the error $\|K^{\text{old}} - K^{\text{new}}\|_\infty + |\omega^{\text{old}} - \omega^{\text{new}}| + \|\mu^{\text{old}} - \mu^{\text{new}}\|_\infty$ between a pair of neighbor steps is less than 10^{-5}. In the case of spatiotemporal model, the terminal conditions are the same except the definition of the error, which is the maximum value among $\|K^{\text{old}} - K^{\text{new}}\|_\infty$, $|\omega^{\text{old}} - \omega^{\text{new}}|$, $\|\mu^{\text{old}} - \mu^{\text{new}}\|_\infty$, $|\sigma^{\text{old}} - \sigma^{\text{new}}|$. The iteration terminates when error is less than 5×10^{-5}.

5.1.1 Comparison Between Parametric Fitting and Non-parametric Fitting

A temporal point process with $U = 10$ categories is simulated. The background rate $\mu_i = 0.1$, $i = 1, 2, \ldots, 10$. The recursive function $h(x) = 2e^{-2x}$ and the triggering function $g(t) = e^{-t}$. And the productivity matrix K is shown in Fig. 4a. Apply the simulation algorithm 3 to generate a temporal point process until $T = 5000$. There are 10,566 cases in this simulated point process. Table 1 and Fig. 5 show the results of the parametric fitting and the non-parametric fitting. In Table 1, the $*$ in iteration means the algorithm terminates because of the reach of maximum number of iteration rather than the reach of the error bound. It can be seen that errors of both fitting algorithms are similar. The non-parametric fitting process can fit K, g, μ, and

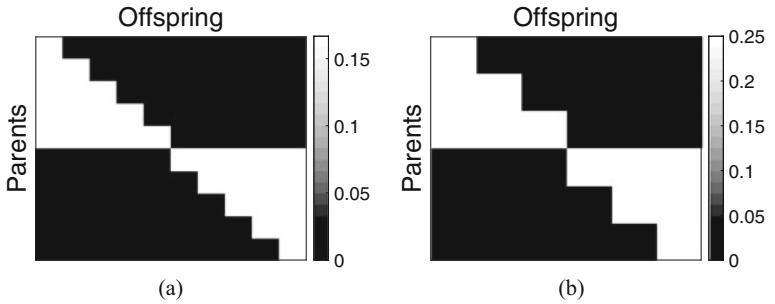


Fig. 4 Ground truth of K matrix. (**a**) $U = 10$. (**b**) $U = 6$

Table 1 The comparison between parametric and non-parametric fitting

	# Iteration	Error μ	Error g	Error K	Error H
Non-parametric	1000*	1.54e−1	1.00e−2	1.72e−1	2.11e−1
Parametric	377	1.58e−1	6.74e−3	1.54e−1	1.54e−1

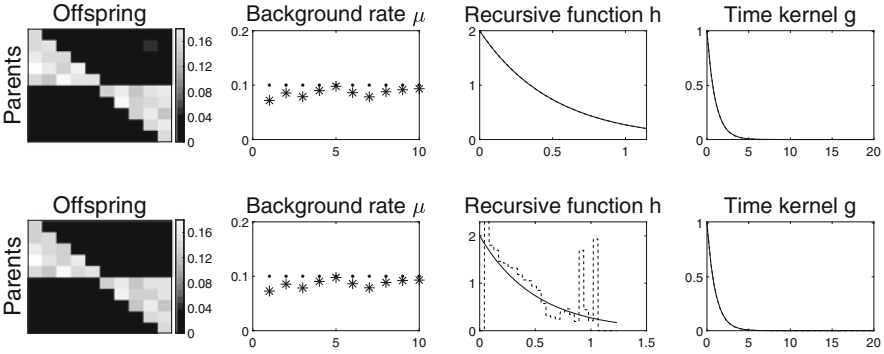


Fig. 5 Comparison of fitting results: The first row are fitted parameters from the non-parametric algorithm, and the second row are fitted parameters from the parametric algorithm. The first column is the productivity matrix K. The second column is the background rate μ, where black dots are the ground truth and asterisks are fitted results. The third and fourth columns are the recursive function h and time-triggering function g, respectively, where black solid lines are the ground truth and black dotted lines are fitted results. In fact, except the non-parametrically fitted h, other fitted results of h and g are almost the same as the ground truth

the productivity number H quite well and can grab the major pattern of function h. But it still has a relatively large error of h in some intervals.

5.1.2 Verification of the Parameter Fitting Algorithm

In this part, different combinations of parameters are used to simulate several point processes and compared with the fitted parameters from the simulated process.

The triggering function and the recursive function are assumed to be parametric. Specially, the temporal-triggering function and the recursive function are the exponential distribution function $g_1(t) = \omega e^{-\omega t}$ and $h(x) = \alpha e^{-\alpha x}$, respectively, and that the spatiotemporal recursive function is the Gaussian distribution function $g_2(x, y) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right)$. The productivity number of the point process is $H(\lambda, i, j) = K_{i,j}h(\lambda)$. Besides, there is a parameter U in experiments, which is the number of sub-processes (categories) of the MRHP.

Two kinds of the productivity matrix K are included in the following experiments, the special one shown in Fig. 4 and the random one. Similarly, for background rate μ, there are the special one $\mu_i = 0.1,\ i = 1, 2, \ldots, U$ and the random one. The way of randomness is to sample each element according to $\mathcal{U}[0, 1]$ (uniform distribution on $[0, 1]$). Then normalize the matrix K and the vector μ that $\sum_{i,j} K^U_{i,j=1} = U/2$ and $\sum_{i=1}^U \mu_i = U/10$.

The terminal conditions are the same as in Sect. 5.1.1. A point process is simulated until $T = 3000$ via Algorithm 3. For spatiotemporal models, spatial locations of the background event are sampled within the range $[0, 5] \times [0, 5]$.

Table 2 shows the results of several experiments with different settings. In the table, N means "not applicable." For temporal models, σ is N and $U = 10$, while for spatiotemporal models, there are values of σ and $U = 6$. S and R under parameters K and μ correspond to the special choice of K or μ and the randomly generated K or μ, respectively.

According to experiments 1–6 about the temporal point processes, the errors about each parameter are stable among different experiment settings. Almost all the relative errors of ω and α are below 0.1. The relative 2-norm errors of K and μ are around 0.2 and 0.1, respectively. The relative errors of productivity number H are around 0.2. According to experiments 7–12 about the spatiotemporal point processes, the relative errors of ω and μ are below 0.1. The relative errors of σ are around 0.3. However, the fitting of matrix K and parameter α about function h performs poorly, some of which are even far from the ground truth. Considering the productivity number $H = K * h$, the parameter fitting algorithm still has a good performance. The relative error of H is around 0.2, which is similar as the temporal cases. Such a result implies that the algorithm cannot fit recursive function h and productivity matrix K separately well, while it still has a good fitting of their product, the productivity number H.

5.1.3 Experiments About Data Sets with Imprecise Time

In some data sets such as COVID-19 data, the exact time of each case is not provided. In this part, some experiments are designed based on simulated data sets to make a comparison of fitted parameters of the original simulated point process, the point process of rounded time values, and the point process with reconstructed time values.

Table 2 Experiments for parameter fitting algorithm

Exp	Parameters						Info		Error					
no.	U	ω	α	σ	K	μ	# Cases	# Iter	ω	α	σ	K	μ	H
Exp 1	10	1	2	N	S	S	6511	381	5.83e−2	6.84e−2	N	1.83e−1	1.51e−1	1.82e−1
Exp 2	10	1.5	1.5	N	S	S	5632	284	5.58e−2	9.74e−2	N	1.75e−1	1.11e−1	1.83e−1
Exp 3	10	1.5	1.5	N	R	S	6016	804	2.31e−2	8.62e−2	N	2.70e−1	9.05e−2	2.53e−1
Exp 4	10	1.5	1.5	N	S	R	5852	396	2.42e−3	1.87e−1	N	2.06e−1	1.23e−1	1.70e−1
Exp 5	10	1.5	1.5	N	R	R	6379	344	1.45e−1	1.06e−1	N	2.22e−1	1.17e−1	2.14e−1
Exp 6	10	1	2	N	R	R	7037	733	5.66e−3	1.03e−1	N	1.92e−1	1.54e−1	1.94e−1
Exp 7	6	1	2	2	S	S	5718	457	3.56e−2	9.67e−1	3.63e−1	25.5	7.17e−2	1.59e−1
Exp 8	6	1	1	1	S	S	2915	969	4.42e−3	7.78	2.00e−1	9.04e−1	6.44e−2	2.31e−1
Exp 9	6	1.5	1.5	1.5	S	S	4164	354	2.00e−2	3.00e−1	2.88e−1	3.34e−1	5.99e−2	1.52e−1
Exp 10	6	1.5	1.5	1.5	S	R	3951	1000*	4.82e−2	1.44	2.86e−1	6.44e−1	8.76e−2	1.69e−1
Exp 11	6	1.5	1.5	1.5	R	S	5072	1000*	7.68e−2	9.44e−1	2.85e−1	14.6	9.12e−2	1.29e−1
Exp 12	6	1.5	1.5	1.5	R	R	4660	601	6.09e−2	3.27	2.81e−1	7.82e−1	9.15e−2	1.40e−1

*In iteration means the algorithm terminates because of the reach of the maximum number of iterations rather than the reach of the error bound

For a certain choice of parameters, the simulation algorithm 3 is applied to generate a temporal point process $X = \{t_i, u_i\}_{i=1}^N$. Let $\bar{t}_i = \lfloor t_i \rfloor$ and define $\overline{X} = \{\bar{t}_i, u_i\}_{i=1}^N$, where $\overline{X}$ is the point process with integer-time values. The reconstruction algorithms 4 and 5 are applied to get a reconstructed point process $X^{\text{recon1}} = \{t_i^{\text{recon1}}, u_i^{\text{recon1}}\}_{i=1}^N$ based on the point process $\overline{X}$. Meanwhile, the random reconstruction method is applied to get a reconstructed point process $X^{\text{recon2}} = \{t_i^{\text{recon2}}, u_i\}_{i=1}^N$, where $t_i^{\text{recon2}} = \bar{t}_i + R_i$ with a uniform random variable R_i on $[0, 1]$. The parametric fitting algorithm 2 is applied to fit parameters of X, $\overline{X}$, X^{recon1}, and X^{recon2} and then compare the fitted parameters with the ground truth.

Table 3 shows the results about experiments on the reconstruction algorithm. Since only temporal processes are considered in this part, parameters about this model are ω, α, K, μ, and H, whose definitions are in Sect. 5.1.2. The errors of these parameters are defined by (37) and (38). There are the special choice and the random choice of triggering matrix K and background rate μ as the notations S and R in Table 3. The way of special and random choices is the same as that in Sect. 5.1.2.

According to Table 3 and Fig. 6, the reconstruction algorithms 4 and 5 and the random reconstruction method can significantly recover fitted results based on the rounded point process. From Fig. 6, the major patterns of matrix K are successfully recovered. In each experiment, the fitted results of the rounded point process are very far away from the ground truth. They even do not include any useful information. This is because the rounded-off process obscures the inner relationship between cases and breaks the smoothness of the original point process. After the reconstruction, parts of the information of the original point process are recovered. From error values in Table 3, the fitted results after reconstruction are worse than those of the original point process, but they are much better than those of the rounded point process. The reconstruction algorithms 4 and 5 have similar performance as the random reconstruction method. Compared with the random reconstruction, Algorithms 4 and 5 are better on the fitting of background rate μ and the productivity number H but are worse on the fitting of time-triggering function g. Meanwhile, the number of iterations needed are similar between these two kinds of reconstructions. It seems that the reconstruction algorithm takes more efforts but performs similar as the trivial random method. However, in Sect. 5.2.1, it is shown that the Algorithms 4 and 5 converge much faster in the point process where cases happen more densely and the process is not a strictly Hawkes process.

5.2 Experiments on Real COVID-19 Data

The data set in this section is the daily reported cases of COVID-19 in California, which is available on the website "Tracking COVID-19 in California" [5]. It provides the number of daily reported cases, daily reported tests, and daily reported deaths of each county in California. Since Los Angeles is the county of most cases

Table 3 Experiments results about the reconstruction algorithm. There are 5 experiments with different choices of parameters. In each experiment, the errors about the original point process, the rounded point process, the reconstructed point process (Recon 1) via Algorithms 4, 5, and the random reconstructed point process (Recon 2) are compared

Exp	Parameters					Info			Error				
no.	U	ω	α	K	μ	Type	# Cases	# Iter	ω	α	K	μ	H
Exp 1	10	1	2	S	S	Original	6308	375	2.77e−2	3.35e−2	1.66e−1	1.17e−1	1.68e−1
						Rounded		1000*	9.17e−1	8.67	2.01	5.20e−1	1.02
						Recon 1		614	1.55e−1	4.58e−2	2.86e−1	2.31e−1	2.95e−1
						Recon 2		648	1.99e−1	2.75e−1	2.60e−1	1.79e−1	2.67e−1
Exp 2	10	1	2	R	R	Original	7004	675	2.90e−2	8.11e−2	2.89e−1	7.99e−2	2.77e−1
						Rounded		1000*	9.60e−1	6.04	8.94e−1	8.99e−1	8.78e−1
						Recon 1		845	2.63e−2	1.78e−1	3.38e−1	1.08e−1	3.13e−1
						Recon 2		843	2.31e−1	2.80e−1	3.30e−1	1.21e−1	2.70e−1
Exp 3	6	1	2	S	S	Original	3915	218	5.31e−2	1.16e−1	1.52e−1	5.42e−2	1.41e−1
						Rounded		628	8.87e−1	6.18	5.84e−1	1.48e−1	7.33e−1
						Recon 1		300	1.53e−1	2.88e−1	2.56e−1	1.51e−1	2.45e−1
						Recon 2		320	2.10e−1	3.08e−1	3.02e−1	1.44e−1	2.38e−1
Exp 4	6	1.5	1.5	S	S	Original	3487	179	2.12e−2	4.71e−2	1.59e−1	7.03e−2	1.41e−1
						Rounded		1000*	8.72e−1	12.8	6.74e−1	3.27e−1	8.18e−1
						Recon 1		202	2.17e−1	2.72e−1	3.90e−1	2.16e−1	3.00e−1
						Recon 2		280	3.06e−1	2.41e−1	3.52e−1	1.99e−1	2.58e−1
Exp 5	6	1	2	R	R	Original	4258	459	5.57e−2	4.00e−1	4.27e−1	1.18e−1	2.02e−1
						Rounded		915	9.40e−1	2.42	5.84e−1	3.78e−1	6.17e−1
						Recon 1		453	1.83e−1	4.23e−1	5.27e−1	2.04e−1	2.97e−1
						Recon 2		261	2.85e−1	5.17e−1	7.26e−1	1.87e−1	2.66e−1

*In iteration means the algorithm terminates because of the reach of the maximum number of iterations rather than the reach of the error bound

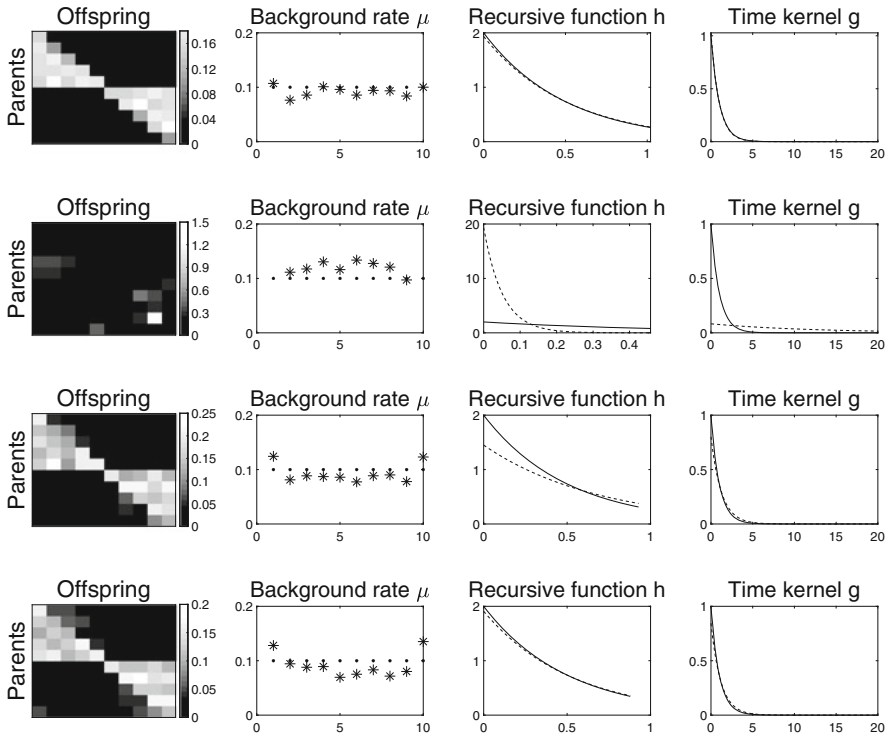


Fig. 6 Experiment results about the reconstruction algorithm. Panels here are corresponding to experiment 1 in Table 3. The first row are fitted results from the original point process X. The second row are fitted results from the rounded point process $\overline{X}$. The third row are fitted results from the reconstructed point process X^{recon1} via reconstruction algorithms 4 and 5. The fourth row are fitted results from the random reconstructed point process X^{recon2}. The four columns are the productivity matrix K, the background rate μ, the recursive function h, and the triggering function g from left to right. In column 2, black dots are the ground truth and asterisks are fitted results. In columns 3 and 4, black solid lines are the ground truth and black dotted lines are fitted results

in CA, daily reported cases of 9 closest counties (include LA) to LA are chosen and considered as a 9-tuple temporal point process. Here the counties chosen are Los Angeles, Orange, Ventura, San Bernardino, Riverside, Kern, San Diego, Santa Barbara, and Tulare. The data set is between 2/1/2020 and 4/11/2021, spanning 435 days. Since the number of cumulative cases is large, a unit size N is applied to rescale the data set, i.e., 1 unit = N cases. Denote $t = 0$ as the date 2/1/2020 and $t = 435$ as the date 4/11/2021. Two problems are raised about the implementation of MRHP on real COVID-19 data:

1. Does the MRHP adequately capture the dynamics of the spread of COVID-19 in California?
2. How accurately can the MRHP forecast future incidence of COVID-19 using past behavior as a guide?

For the first problem, experiments are designed about different time windows in $[0, 435]$ of the original data set. For each subset, the parameter fitting algorithm is applied to estimate the triggering matrix K, the time-triggering function g, the recursive function h, and the background rate μ, where the time-triggering function g and the recursive function h are assumed parametric. Several point processes are simulated based on these fitted parameters using the simulation algorithm 3. The simulated point processes are compared with the original one to see if they are similar enough.

For the second problem, experiments are also based on different time intervals $[t_1, t_2] \subset [0, 435]$. Let X be the point process corresponding to the time interval $[t_1, t_2]$. Apply parameter fitting algorithms 2 to estimate the parameters about point process X. Setting X as the history process, using fitted parameters, simulate the point process in the time interval $[t_2, t_2 + t]$ and compare the simulated results with the ground truth.

5.2.1 Model Validation

For each point process X_i, the parameters of MRHP are estimated and applied to simulate M point processes $X_i^{(j)},\ j = 1, 2, \ldots, M$. Define the frequency vector of $X_i^{(j)}$ as $f_{i,j} = (f_{i,1}^{(j)}, f_{i,2}^{(j)}, \ldots, f_{i,n_i}^{(j)})$, where $n_i = t_{i,2} - t_{i,1}$ and $f_{i,k}^{(j)}$ is the number of cases of $X_i^{(j)}$ in the interval $[t_{i,1}+k-1, t_{i,1}+k)$. Similarly, the frequency vector of X_i is f_i. For $p \in (0, 1)$, define the $F_{i,k}^p$ as the set $\{f_{i,k}^{(j)},\ j = 1, 2, \ldots, M\}$ dropping the $\frac{(1-p)M}{2}$ greatest values and the $\frac{(1-p)M}{2}$ least values. Roughly speaking, if $p = 0.9$, then the set $F_{i,k}^p$ is the set of values $f_{i,k}^{(1)}, f_{i,k}^{(2)}, \ldots, f_{i,k}^{(M)}$ without the top 5% and the bottom 5%. Further, define the **p-confidence range** of simulated frequency on the interval $[t_{i,1} + k - 1, t_{i,1} + k)$ as $I_{i,k}^p = [\min F_{i,k}^p, \max F_{i,k}^p]$. Then the simulation error is defined by

$$\text{error}(f_i) = \sum_{k=1,2,\ldots,n_i} \frac{\mathbb{1}_{f_{i,k} \notin I_{i,k}^p}}{n_i}. \tag{39}$$

The raw data set of daily reported cases is shown in Fig. 7. It should be noticed that there is a significant increasing of the number of reported cases after $T = 300$. Considering two groups of time intervals, the first group includes 4 intervals $I_1 = [50, 250],\ I_2 = [50, 280],\ I_3 = [50, 300],\ I_4 = [100, 300]$, which do not include the big explosion after $T = 300$ and the second group includes 6 intervals $I_5 = [150, 350],\ I_6 = [200, 400],\ I_7 = [100, 350],\ I_8 = [150, 400],\ I_9 = [50, 360],\ I_{10} = [50, 400]$. The corresponding unit size $N_1 = 120$, $N_2 = 200$, $N_3 = N_4 = 160$, $N_5 = N_7 = 240$, $N_9 = 280$, and $N_6 = N_8 = N_{10} = 320$. When calculating the frequency ranges, the number of simulations $M = 300$.

The parameter fitting algorithm terminates if the number of iterations reaches its maximum value, 3000, the difference of likelihood function (27) between a pair of

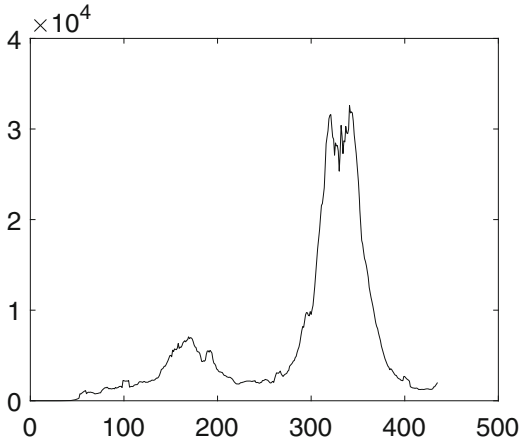


Fig. 7 Smoothed data of daily reported cases

neighbor steps is less than 10^{-6}, or the error $\|K^{\text{old}} - K^{\text{new}}\|_\infty + |\omega^{\text{old}} - \omega^{\text{new}}| + \|\mu^{\text{old}} - \mu^{\text{new}}\|_\infty$ between a pair of neighbor steps is less than 5×10^{-5}.

The daily reported cases of nine counties within time interval I_i form a point process X_i for $i = 1, 2, \ldots, 10$. It takes two steps to preprocess the raw point process X_i:

1. **Smoothing:** Replace the raw daily reported cases by the mean value of the closest 7 days.
2. **Reconstruction:** Apply the reconstruction algorithms on the smoothed point process.

First, the number of iterations are compared between different reconstruction algorithms. The subinterval I_1, I_4, I_5, I_6, I_9, I_{10} and their corresponding point processes X_i, $i = 1, 4, 5, 6, 9, 10$ are chosen to make the test. Table 4 shows the comparison results. It can be seen that the fitting takes much less iterations on the point process reconstructed by the Algorithms 4 and 5. Taking the time interval $[50, 400]$ as an example, there are 7811 (rescaled) cases in 351 days, while the synthetic data in the previous section has around 6000 cases in the time interval $[0, 3000]$. The real data are much more dense. Besides, the real data may not be a standard Hawkes process. In such a circumstance, the reconstruction algorithm outperforms. As a consequence, the reconstruction algorithms 4 and 5 are applied in the preprocess step.

The preprocessing steps output point processes $\widetilde{X}_i$, $i = 1, 2, \ldots, 10$. Parameters are fitted via parametric fitting algorithm 2 under the assumption that the recursive function h and the triggering function g are both exponential distribution function. Figure 8 shows the experiment results. Table 5 shows the error values of each experiment. From the panels about simulation curves in Fig. 8, most of the ground truth curves are covered by the simulation range. Meanwhile, from Table 5, there are 7 out of 10 experiments with the simulation error error(f) less than 20%.

Table 4 Comparison of the number of iterations between two different reconstruction algorithms. Recon 1 refers to the reconstruction algorithms 4 and 5. Recon 2 refers to the naive random reconstruction algorithm. The notation $*$ means the iterations terminate because of reaching the maximum number of iterations rather than reaching the error requirements

Exp no.	Start T	End T	N	# Iteration Recon 1	# Iteration Recon 2
Exp 1	50	250	120	1071	1764
Exp 2	100	300	160	2037	2296
Exp 3	150	350	240	2582	2532
Exp 4	200	400	320	1168	1698
Exp 5	50	360	280	1031	3000*
Exp 6	50	400	320	866	1541

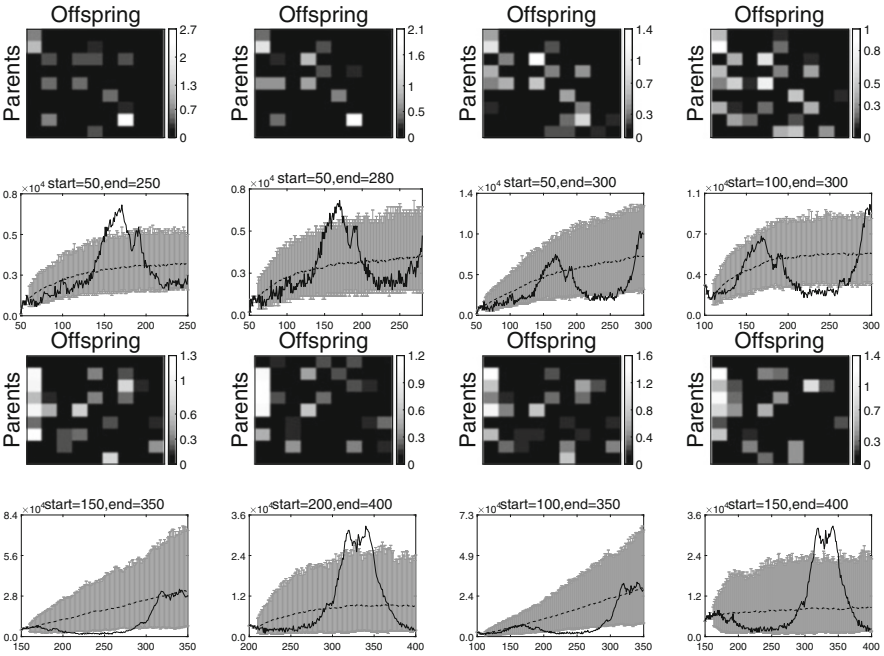


Fig. 8 Row 1,2 are experiment results about point process $\widetilde{X}_i,\ i = 1, 2, 3, 4$ in group 1. Row 3,4 are experiment results about point process $\widetilde{X}_i,\ i = 5, 6, 7, 8$ in group 2. Panels in row 1,3 are the productivity matrices K, and the panels in row 2,4 are the comparisons between the ground truth frequency (daily reported cases) of X_i and the simulated frequencies. For each panel about the K matrix, the rows from top to bottom (also the columns from left to right) correspond to county Los Angeles, Orange, Ventura, San Bernardino, Riverside, Kern, San Diego, Santa Barbara, and Tulare, respectively. In each panel about the simulation curve, the black solid curve is the ground truth; the gray region is the 90% confidence interval of simulation, and the black dotted line is the average of simulation process, where x-axis is the time and y-axis is the number of cases

Table 5 Error results of experiments on 2 groups. There are 4 experiments with different choices of parameters. In each experiment, the simulation error (39) is applied to evaluate the performance of parameter fitting and simulation

Group no.	Exp no.	Start T	End T	N	Error f
Group 1	Exp 1	50	250	120	1.64e−1
	Exp 2	50	280	200	8.23e−2
	Exp 3	50	300	160	1.63e−1
	Exp 4	100	300	160	2.94e−1
Group 2	Exp 5	150	350	240	4.08e−1
	Exp 6	200	400	320	1.64e−1
	Exp 7	100	350	240	3.23e−1
	Exp 8	150	400	320	1.51e−1
	Exp 9	50	360	280	8.68e−2
	Exp 10	50	400	320	1.31e−1

5.2.2 Prediction Based on MRHP and Historical Information

Table 6 shows the results from 12 different experiments, and Figs. 9 and 10 visualize the corresponding predictions. For each experiment, a short-term prediction for the following 15 days and a long-term prediction for the following 50 days are made based on fitted parameters. The error metrics error(f) in Sect. 5.2.1 are applied to evaluate the performance of prediction, whose definitions are the same as that in Sect. 5.2.1.

It can be seen that short-term predictions are fairly good, while most of the long-term predictions are not acceptable. Especially when predicting on some sharp increasing/decreasing parts, the predicted curves are far away from the ground truth. The sharp increasing and decreasing might be because of some external factors such as the change of policy or the promotion of vaccine. Since the MRHP is a self-excited model, these external factors will affect the performance of prediction.

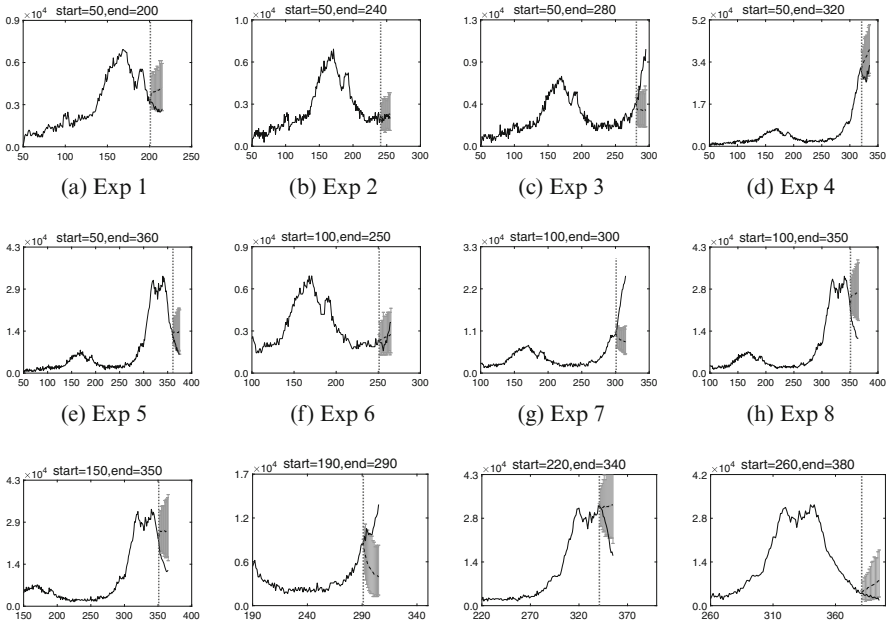
6 Conclusion

In this chapter, the multivariate recursive Hawkes process (MRHP) is introduced. Several properties of it are derived, and the performance on COVID-19 data is analyzed based on experiments. The existence of a counting measure with a MRHP intensity is proved. Formulas of mean and variance about the counting process are derived. It is shown how the MRHP can be fitted to data using an expectation–maximization (EM) algorithm and provided an algorithm for simulating realizations of the process given its parameters. Several experiments on synthetic data were used to verify the simulation and parameter fitting algorithms.

In COVID-19 data sets, event timestamps are typically unavailable, and counts are instead provided daily. This binning of event times can impact the performance of the parameter estimation algorithm, and such cases might lead to incorrect interpretations of model parameters (e.g., the reproduction number). To address this

Table 6 Experiments about the prediction performances. There are 12 experiments with different start dates and end dates. SE means short-term error, and LE means long-term error

Exp no.	Start T	End T	N	SE	LE	Exp no.	Start T	End T	N	SE	LE
Exp 1	50	200	100	2.00e−1	7.20e−1	Exp 7	100	300	200	8.00e−1	9.40e−1
Exp 2	50	240	150	0	1.60e−1	Exp 8	100	350	250	9.33e−1	9.80e−1
Exp 3	50	280	200	6.67e−1	9.00e−1	Exp 9	150	350	250	8.00e−1	9.40e−1
Exp 4	50	320	240	3.33e−1	7.20e−1	Exp 10	190	290	200	6.67e−1	9.00e−1
Exp 5	50	360	280	2.00e−1	7.80e−1	Exp 11	220	340	250	2.67e−1	7.80e−1
Exp 6	100	250	150	0	3.00e−1	Exp 12	260	380	280	1.33e−1	6.80e−1



(a) Exp 1 (b) Exp 2 (c) Exp 3 (d) Exp 4

(e) Exp 5 (f) Exp 6 (g) Exp 7 (h) Exp 8

(i) Exp 9 (j) Exp 10 (k) Exp 11 (l) Exp 12

Fig. 9 Experiments about short-term predictions for the following 15 days. In each panel, there is a dotted vertical line that marks the point between the historical information and the prediction. The solid black curve is the ground truth, the gray region is the 90% confidence region of prediction, and the black dotted curve is the average of prediction. The start date, end date, and the unit size N of each experiment are in the title of each panel and can also be found in Table 6. In each panel, x-axis is the time and y-axis is the number of cases. (**a**) Exp 1. (**b**) Exp 2. (**c**) Exp 3. (**d**) Exp 4. (**e**) Exp 5. (**f**) Exp 6. (**g**) Exp 7. (**h**) Exp 8. (**i**) Exp 9. (**j**) Exp 10. (**k**) Exp 11. (**l**) Exp 12

issue, two possible strategies are investigated to deal with the problem of coarse-grained event times. A shape-preserving piecewise cubic interpolation method is applied, which can reconstruct the timestamps from coarse-grained event times during model fitting. Through experiments using synthetic data, it is observed that the reconstruction algorithm can improve the performance of parameter fitting.

Several experiments on real COVID-19 data are conducted to infer the key parameters such as the productivity (reproduction) matrix K under the MRHP model. However, it should be noticed that the MRHP model had difficulty in medium- and long-term forecasting of COVID-19 surges. There are a number of exogenous factors not included in the model that may be necessary to improve longer-term forecasts. This is a significant challenge for infectious disease forecasting models in general, as SEIR-type models also have such difficulties. Predicting when surges will occur, what shape they will take, and when they will subside should be a direction for future research, either by improving model performance or developing insights into why such a problem may be ill-posed.

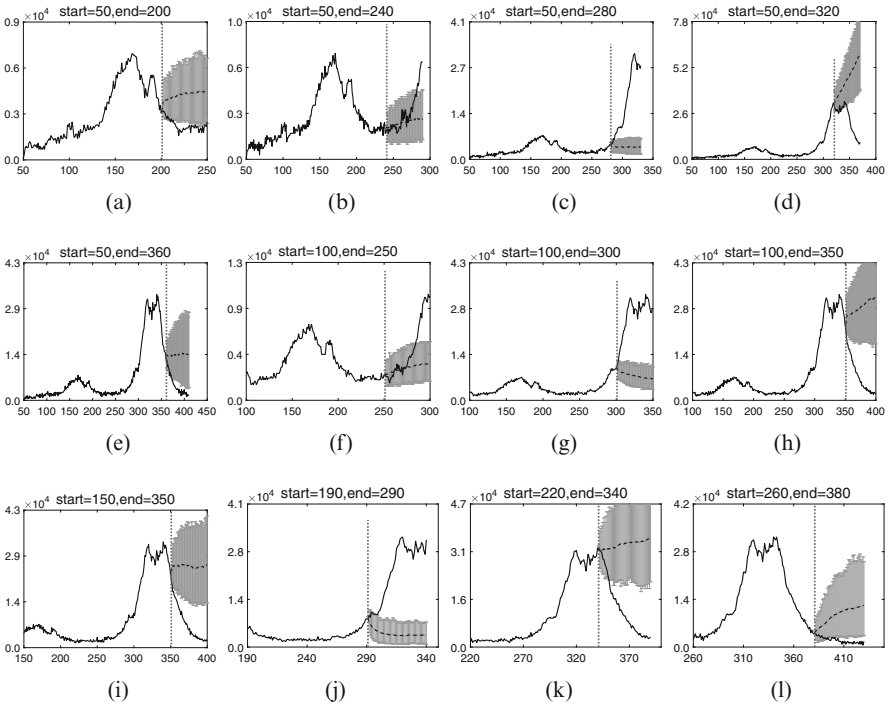


Fig. 10 Experiments about long-term predictions for the following 50 days. In each panel, there is a dotted vertical line that marks the point between the historical information and the prediction. The solid black curve is the ground truth, the gray region is the 90% confidence region of prediction, and the black dotted curve is the average of prediction. The start date, end date, and the unit size N of each experiment are in the title of each panel and can also be found in Table 6. In each panel, x-axis is the time and y-axis is the number of cases. (**a**) Exp 1. (**b**) Exp 2. (**c**) Exp 3. (**d**) Exp 4. (**e**) Exp 5. (**f**) Exp 6. (**g**) Exp 7. (**h**) Exp 8. (**i**) Exp 9. (**j**) Exp 10. (**k**) Exp 11. (**l**) Exp 12

Acknowledgment This work is supported by NSF grants DMS-2027277, DMS-1737770, DMS-2124313, and DMS-2027438 and Simons Foundation Math + X Investigator Award # 510776.

References

1. H. Akima, A new method of interpolation and smooth curve fitting based on local procedures. J. ACM **17**(4), 589–602 (1970)
2. H. Akima, A method of bivariate interpolation and smooth surface fitting for irregularly distributed data points. ACM Trans. Math. Softw. **4**(2), 148–159 (1978)
3. A.L. Bertozzi, E. Franco, G. Mohler, M.B. Short, D. Sledge, The challenges of modeling and forecasting the spread of COVID-19. Proc. Natl. Acad. Sci. **117**(29), 16732–16738 (2020)
4. T. Björk, An introduction to point processes from a martingale point of view. Lecture note, KTH, 2011
5. California covid-19 data. https://covid19.ca.gov/state-dashboard/

6. S. Cauchemez, P. Nouvellet, A. Cori, T. Jombart, T. Garske, H. Clapham et al., Unraveling the drivers of MERS-CoV transmission. Proc. Natl. Acad. Sci. **113**(32), 9081–9086 (2016)
7. Z. Chen, A. Dassios, V. Kuan, J.W. Lim, Y. Qu, B. Surya, H. Zhao, A two-phase dynamic contagion model for COVID-19. Results Phys. **26**, 104264 (2021)
8. W.H. Chiang, X. Liu, G. Mohler, Hawkes process modeling of COVID-19 with mobility leading indicators and spatial covariates. Int. J. Forecast. **38**(2), 505–520 (2022)
9. D.J. Daley, D. Vere-Jones, *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods* (Springer, New York, 2003)
10. A. Dassios, H. Zhao, A dynamic contagion process. Adv. Appl. Probab. **43**(3), 814–846 (2011)
11. S. Delattre, N. Fournier, M. Hoffmann, Hawkes processes on large networks. Ann. Appl. Probab. **26**(1), 216–261 (2016)
12. P. Diggle, B. Rowlingson, T.L. Su, Point process methodology for on-line spatio-temporal disease surveillance. Environ. Off. J. Int. Environmetrics Soc. **16**(5), 423–434 (2005)
13. E.W. Fox, M.B. Short, F.P. Schoenberg, K.D. Coronges, A.L. Bertozzi, Modeling e-mail networks and inferring leadership using self-exciting point processes. J. Am. Stat. Assoc. **111**(514), 564–584 (2016)
14. E.W. Fox, F.P. Schoenberg, J.S. Gordon, Spatially. inhomogeneous background rate estimators and uncertainty quantification for nonparametric Hawkes point process models of earthquake occurrences. Ann. Appl. Stat. **10**(3), 1725–1756 (2016)
15. G.H. Hardy, J.E. Littlewood, G. Pólya, *Inequalities* (Cambridge University Press, Cambridge, 1952)
16. A.G. Hawkes, Spectra of some self-exciting and mutually exciting point processes. Biometrika **58**(1), 83–90 (1971)
17. A.G. Hawkes, D. Oakes, A cluster process representation of a self-exciting process. J. Appl. Probab. **11**(3), 493–503 (1974)
18. A. Kaplan, J. Park, F.P. Schoenberg, Nonparametric estimation of recursive point processes with application to mumps in Pennsylvania (2020)
19. M. Kim, D. Paini, R. Jurdak, Modeling stochastic processes in disease spread across a heterogeneous social system. Proc. Natl. Acad. Sci. **116**(2), 401–406 (2019)
20. L. Lesage, A Hawkes process to make aware people of the severity of COVID-19 outbreak: application to cases in France. Doctoral dissertation, Université de Lorraine; University of Luxembourg, 2020
21. T.J. Liniger, Multivariate Hawkes processes. Doctoral dissertation, ETH Zurich, 2009
22. J.O. Lloyd-Smith, S. Funk, A.R. McLean, S. Riley, J.L. Wood, Nine challenges in modelling the emergence of novel pathogens. Epidemics **10**, 35–39 (2015)
23. D. Marsan, O. Lengline, Extending earthquakes' reach through cascading. Science **319**(5866), 1076–1079 (2008)
24. S. Meyer, L. Held, Power-law models for infectious disease spread. Ann. Appl. Stat. **8**(3), 1612–1639 (2014)
25. Y. Ogata, Space-time point-process models for earthquake occurrences. Ann. Inst. Stat. Math. **50**(2), 379–402 (1998)
26. A. Reinhart, A review of self-exciting spatio-temporal point processes and their applications. Stat. Sci. **33**(3), 299–318 (2018)
27. F.P. Schoenberg, A note on the consistent estimation of spatial-temporal point process parameters. Stat. Sinica **26**, 861–879 (2016)
28. F.P. Schoenberg, M. Hoffmann, R.J. Harrigan, A recursive point process model for infectious diseases. Ann. Inst. Stat. Math. **71**(5), 1271–1287 (2019)
29. H.J.T. Unwin, I. Routledge, S. Flaxman, M.A. Rizoiu, S. Lai, J. Cohen et al., Using Hawkes processes to model imported and local malaria cases in near-elimination settings. PLoS Comput. Biol. **17**(4), e1008830 (2021)
30. A. Veen, F.P. Schoenberg, Estimation of space–time branching process models in seismology using an EM-type algorithm. J. Am. Stat. Assoc. **103**(482), 614–624 (2008)
31. H. Xu, M. Farajtabar, H. Zha, Learning Granger causality for Hawkes processes, in *International Conference on Machine Learning*. PMLR, June 2016, pp. 1717–1726

32. A.S. Yang, Modeling the transmission dynamics of pertussis using recursive point process and SEIR model. Doctoral dissertation, UCLA, 2019
33. B. Yuan, H. Li, A.L. Bertozzi, P.J. Brantingham, M.A. Porter, Multivariate spatiotemporal Hawkes processes and network reconstruction. SIAM J. Math. Data Sci. **1**(2), 356–382 (2019)
34. J. Zhuang, Weighted likelihood estimators for point processes. Spat. Stat. **14**, 166–178 (2015)
35. J. Zhuang, Y. Ogata, D. Vere-Jones, Analyzing earthquake clustering features by using stochastic reconstruction. J. Geophys. Res. Solid Earth **109**(B5) (2004). https://doi.org/10.1029/2003JB002879

Multiscale Aspects of Virus Dynamics

Franco Flandoli, Eleonora La Fauci, and Martina Riva

1 Introduction

The diffusion of a virus in a human population is a key example of a multiscale system. The virus operates inside the human body at cellular level with consequences also at systemic level. These two aspects already concern very different scales that give rise to heterogeneous behaviors and represent a complex system in itself with different outputs and a large degree of uncertainty.

Then we move to the interactions between humans, which are the reason for the virus diffusion. These interactions concern a much larger scale and give rise to a great degree of complexity and source of uncertainty. Then one can summarize the information about the virus diffusion in a population by means of macroscopic variables, like the well-known S, I, R in the homonymous model that we shall review below. And different scales may also enter here, depending whether these variables are computed for a single town, a nation, or the entire world.

Few other examples in natural sciences have such amount of scales interlaced in a complex manner. Examples like turbulence, although extremely difficult and still partially open, may benefit of properties like some degree of self-similarity at different scales; it is not so for virus diffusion. A somewhat similar example is the understanding of weather-climate dynamics, where world-wide phenomena should be understood which, however, are influenced by local inputs like relatively small convective cells and human activity.

Our aim in this chapter is discussing the duality between compartmental models (which we could call macroscopic models) and individual-based models (microscopic with the respect to the compartmental ones, although not yet truly

F. Flandoli (✉) · E. La Fauci · M. Riva
Scuola Normale Superiore, Pisa, Italy
e-mail: franco.flandoli@sns.it

N. Bellomo, M. A. J. Chaplain (eds.), *Predicting Pandemics in a Globally Connected World, Volume 1*, Modeling and Simulation in Science, Engineering and Technology, https://doi.org/10.1007/978-3-030-96562-4_6

microscopic in terms of virus interaction with human cells). This duality is only one aspect of the possible multiscale features of virus dynamics, but we prefer to focus on one aspect in order to enter in some technical details.

Limited to this duality, indeed, we insist on one aspect which is not always properly taken into account in the literature on this topic: a correct quantification of the uncertainty. We distinguish between two sources of uncertainty called epistemic and empirical, and show where, in the models, they appear and how we should properly quantify them. Usually the quantification of uncertainty is a frustrating exercise since there is no data to check whether it has been done correctly (often we deal with a phenomenon which happens only once, like in climate studies). But in the case of COVID-19 pandemic, due to the hard lockdown experienced in Spring 2020, the behavior of certain towns of similar kind may be seen as a repeated experiment under similar and independent conditions. This allows us to compare this experiment with different quantification rules and see advantages and drawbacks of them.

1.1 On the Biology of the Virus

Before we enter Mathematics it may be useful to recall a few features of COVID-19 that we shall meet in the models. We refer to the report [8], which is bimonthly updated, and to the papers [4] and [9] for much wider accounts. A brief description of the biology of the virus is delivered in the following:

- *The virus*: Coronavirus disease 2019 (COVID-19) is initiated by the infection of the SARS-CoV-2 virus, a coronavirus.
- *The contagion*: The contagion by the SARS-CoV-2 occurs when breathing respiratory droplets, up to 1 millimeter across, released by an infected person via coughing, sneezing, or speaking.
- *Entry into human cells*: The large Spike protein (S) forms a sort of crown on the surface of the viral particles. Its receptor-binding domain interacts with high affinity with angiotensin-converting enzyme 2 (ACE2) receptors on the surface of host cells. After the binding, two host cell proteases (Furin and TMPRLRS) cleave spike proteins and their exposed fusion peptides fuse the virus membrane with the membrane of the host cells. The virus RNA enters cells of the upper and lower respiratory tract, and it is translated into viral proteins. The cell dies releasing millions of new viruses that infect other cells and other individuals.
- *Immune system responses*: Cellular and humoral innate immunity represents the first line of resistance which takes care of most encounters with infectious agents. When the virus manages to overcome these barriers, a rapid release of danger signals activates the reaction of the host immunity.
- *Immunity*: There is evidence that symptomatic COVID-19 elicits immunological memory and resistance to reinfection. Based on SARS, one can expect immunological memory to last 2–3 years.

Moreover, we want to focus on two aspects that we will be crucial in the following, as they represent important risk factors in the spreading of the illness.

- *Stages of the incubation period*: The incubation period, defined as the time between exposure to the virus and symptom onset, can be divided into two stages. The first one in which the amount of viruses is still small and the individual cannot transmit the virus to others; and the second stage, still of incubation, in which the individual is infectious because, as the virus multiplies, he/she may shed copious amounts of it. Both are pre-symptomatic stages; hence, the individual has no sign of being infectious: the immune system prepares its answer but still does not act so strongly. Needless to say, the second stage is very dangerous for other individuals.
- *Unknown patients*: There are several studies, such as [12, 18, 25], that show that a large proportion of the population has SARS-CoV-2 infections that do not result in COVID-19 symptoms or result in mild symptoms that go unnoticed. Therefore, there is a large proportion of infected individuals who remain unknown to the National Health Systems and who are not isolated. It is clear that this kind of patient is highly dangerous for other individuals, as they may act as if they are not infectious.

1.2 Modeling the Complexity of COVID-19

Covid-19 massive diffusion all over the world motivated the development and investigation of several mathematical models. Many of them are based on differential equations, ordinary or partial, deterministic or stochastic, or based on other stochastic processes and dynamical systems. Among those that use differential equations we cite [17, 29, 31]. All of these papers are based on the fundamental SIR model introduced by Kermack and McKendrick in [21] in 1927, who idealized the phenomenon in few "compartments"; but still in its simplicity indicates basic logical building blocks which are present in much more refined models. Generalizations and additional complexity due to spatial structure have been introduced later on; let us mention, for instance, [2, 3, 5–7, 13, 22] and extensive treatises like [1, 14, 26]. The list of contributions is very large, see the references in [4] and the examples of multiscale models presented there. All works try to approach some aspects of the multiscale complexity. For instance, viral load and the response of the immune system are included by Perelson [27]. Another key aspect is the network structure between individuals or communities that can be deterministic or stochastic; some examples can be found in [10, 20, 28, 30]; see also [1, 11, 16, 24] for other approaches. With many differences, the methodology described in the present chapter looks similar in spirit to what is done in [4]. Among the peculiarities, let us mention the detailed treatment of the randomness which pervades all steps of our modeling and represents one of main successes in comparison with real data.

2 Epistemic and Empirical Uncertainties in Compartmental and Individual-Based Models

2.1 SIR Model

Recall the equations of SIR model

$$\frac{dS(t)}{dt} = -\lambda \frac{I(t)}{N} S(t),$$
$$\frac{dI(t)}{dt} = \lambda \frac{I(t)}{N} S(t) - \frac{1}{\tau} I(t),$$
$$\frac{dR(t)}{dt} = \frac{1}{\tau} I(t),$$

where N is the total number of individuals, $S(t)$ the number of those which are susceptible, $I(t)$ the number of infected, and $R(t)$ the number of people restored from illness. This is the simplest example of the so-called compartmental models; more elaborated variants include additional compartments, namely categories of individuals (like a subdivision of infected subjects into symptomatic and asymptomatic). In the sequel we discuss the original SIR model above but the arguments we develop apply to all *deterministic* compartmental models (if they are stochastic, the discussion may be modified case by case).

It is known that this simple model or some of its variants are able to fit epidemic curves very well. However, it is different fitting than predicting. In order to make predictions it is necessary to choose in advance relatively precise values of the parameters λ and τ and the solutions depend a lot on these values. The parameter τ is easier, its meaning is the average duration of the illness and thus it can be estimated quite early and remains relatively stable under different conditions. But λ incorporates the effect of so many factors that it is almost impossible to decide its value before having already an experimental curve. Hence it is very difficult to use the SIR model to make projections and decide which policies will be more successful, if these policies correspond to new regimes not experienced before. Less ambitiously, if we just want to explore the qualitative dependence of the projections on different social measures, it is difficult to use SIR because the effect of too many factors is accumulated in the value of the single parameter λ.

Recall the meaning of the two main different concepts of uncertainty called epistemic and empirical. The *epistemic uncertainty* refers to any uncertainty related to the model: algebraic structure of the nonlinearities, variables treated by the model, and finally the numerical values of the parameters. In the case of compartmental models it may refer to the variables considered (e.g., only the infected as a total or dividing them as symptomatic and asymptomatic ones), but let us accept that the elementary SIR model is already a very good approximation or reality. The lack of knowledge about the value of the parameter λ is the major source of epistemic uncertainty in the SIR model.

The *empirical uncertainty* on the contrary refers to the intrinsic variability of a phenomenon. In the case of Covid it comes from many aspects (here the multiscale nature of the problem is particularly relevant). Two of them, of very different nature, are: (1) when an infected and a susceptible individuals meet, not necessarily the susceptible becomes infected, it depends on several unpredictable circumstances; (2) the duration of the infective period for a person is random, it varies in the population, as well as the duration of the incubation period; and the duration of these periods affects the spread of the virus.

Going back to SIR model with the aid of this conceptualization of the uncertainties, we may say that epistemic uncertainty, for SIR, is mainly due to the parameter λ; but empirical uncertainty is completely absent in the model: nothing corresponds to it (as it is λ for the epistemic uncertainty) and no way to quantify the empirical uncertainty is provided by the model.

SIR is a model for global quantities, the total number of individuals with this or that feature. Behind this there are the individuals with their particular story. Models that take into account the single individuals as specific units of the models, with certain variables for each one of them, will be called *individual-based*. The global quantities of the SIR model, at least at an heuristic level, correspond to averages of the individual-based models (there are theorems formalizing this link, see below). The true number of infected, in an individual model, is a stochastic process, with a variability, an uncertainty with respect to the average behavior. Therefore it is natural to expect that individual-based models capture the empirical uncertainty. In Sect. 3.2 we discuss this topic and, in particular, we show that the estimate of intrinsic variability given by individual-based models is very realistic, while shortcuts based on the Central Limit Theorem, immediately applicable to SIR itself, lead to wrong results.

Less clear is whether individual-based models may help to reduce the epistemic uncertainty: a priori they have more parameters, since there are more details and more variables. However, there is a chance that the parameters of the individual-based model are more elementary and unique, less composite: each one corresponding to some specific feature, not summarizing the effect of many inputs as it is for the coefficient λ of the SIR model. We shall see below that the answer to the question whether an individual-based model is an improvement with respect to compartmental model from the viewpoint of the epistemic uncertainty is difficult and may depend on the model. Roughly speaking, to anticipate what we shall see, the most elementary individual-based model behind SIR does not improve, λ remains its major uncontrolled parameter. But we shall discuss a variant of the elementary model, based on more refined multiscale information about the virus diffusion and the development of the infectiousness of the individuals, such that λ is somewhat decomposed in more focused and elementary sub-parameters, having a more direct interpretation; see Sect. 3.1.

2.2 Individual-Based Interpretation of λ

Let us discuss here an individual-based interpretation of λ which may help to get more information on this parameter.

Assume we analyze a community of N subjects, which for simplicity is assumed to be closed to exchanges with the external world. Each subject $i = 1, \dots, N$ has a "state" $X^i(t)$ which may change in time. The state, in this simple model take only the following three possible values:

$$X^i(t) \in \{S, I, R\}.$$

In a short time interval $[t, t + \Delta t]$ an individual i such that $X^i(t) = S$ could jump to the state $X^i(t + \Delta t) = I$. The probability of this jump, for very short time, has the form

$$\mathbb{P}\left(X^i(t + \Delta t) = I | X^i(t) = S\right) \sim \lambda \frac{I(t)}{N} \Delta t, \tag{1}$$

where the notation $\sim$ stands for the fact that the rule is true only in the limit as $\Delta t \to 0$, hence it should be formulated as

$$\lim_{\Delta t \to 0} \mathbb{P}\left(X^i(t + \Delta t) = I | X^i(t) = S\right) / \Delta t = \lambda \frac{I(t)}{N},$$

but we shall omit this more precise formulation in the sequel. Here $I(t)$ is the number of individuals that, at time t, are of class I:

$$I(t) = \sum_{j=1}^{N} 1_{\{X^j(t) = I\}}. \tag{2}$$

The previous prescription is not properly formulated, as we see from the previous rule. Indeed, for simplicity we were tempted to write the conditional probability that individual i jumps from $X^i(t) = S$ to $X^i(t + \Delta t) = I$ but the rule depends on the state at time t of all individuals, not only on the state of individual i, hence it must be a conditional probability to such more complex state. We should write

$$\mathbb{P}\left(X^i(t + \Delta t) = I | \left\{X^j(t)\right\}_{j=1,\dots,N} = \left\{x_j\right\}_{j=1,\dots,N}\right) \sim \lambda \frac{\sum_{j=1}^{N} 1_{\{x_j = I\}}}{N} \Delta t,$$

if $x_i = S$. However, since this more rigorous formulation is less intuitive, we continue to write the first one; the reader may easily correct the formalism.

The model just introduced, complemented by easier rules of passage from I to R

$$\mathbb{P}\left(X^i\left(t+\Delta t\right)=R|X^i\left(t\right)=I\right)\sim\frac{\Delta t}{\tau}$$

(here the formalism is correct), can be proved rigorously to converge to the SIR model, in the following mean field sense. First, let us formulate SIR system of differential equations for the fractions $s\left(t\right)=\frac{S(t)}{N}$, $i\left(t\right)=\frac{I(t)}{N}$, $r\left(t\right)=\frac{R(t)}{N}$, so that the parameter N does not appear anymore:

$$\begin{aligned}\frac{ds\left(t\right)}{dt}&=-\lambda i\left(t\right)s\left(t\right),\\ \frac{di\left(t\right)}{dt}&=\lambda i\left(t\right)s\left(t\right)-\frac{1}{\tau}i\left(t\right),\\ \frac{dr\left(t\right)}{dt}&=\frac{1}{\tau}i\left(t\right).\end{aligned}$$

Then introduce the empirical fractions associated with the individual-based model above:

$$\begin{aligned}s_N\left(t\right)&=\frac{1}{N}\sum_{j=1}^{N}1_{\{X^j(t)=S\}},\\ i_N\left(t\right)&=\frac{1}{N}\sum_{j=1}^{N}1_{\{X^j(t)=I\}},\\ r_N\left(t\right)&=\frac{1}{N}\sum_{j=1}^{N}1_{\{X^j(t)=R\}}.\end{aligned}$$

Then it can be proved that the stochastic process $\left(s_N\left(t\right),i_N\left(t\right),r_N\left(t\right)\right)$ converges in probability, uniformly in time on compact time intervals, to the deterministic function $\left(s\left(t\right),i\left(t\right),r\left(t\right)\right)$. This holds true under proper assumptions on the initial condition; for instance, the following classical one: given a triple of non-negative numbers $\left(s\left(0\right),i\left(0\right),r\left(0\right)\right)$ such that $s\left(0\right)+i\left(0\right)+r\left(0\right)=1$, we assume that the random variables $\left\{X^i\left(0\right)\right\}_{i=1,\ldots,N}$ are independent, identically distributed, with

$$\begin{aligned}\mathbb{P}\left(X^i\left(0\right)=S\right)&=s\left(0\right),\\ \mathbb{P}\left(X^i\left(0\right)=I\right)&=i\left(0\right),\\ \mathbb{P}\left(X^i\left(0\right)=R\right)&=r\left(0\right).\end{aligned}\tag{3}$$

This result can be complemented by a study of fluctuation which, roughly speaking, tells us that the fluctuations behave like

$$\sum_{j=1}^{N} 1_{\{X^i(t)=I\}} - Ni\,(t) \sim C_I\,(t)\,\sqrt{N},$$

for a suitable function $C_I\,(t)$. In Sect. 3.2 below we discuss this rule and the value of $C_I\,(t)$.

Let us examine the relation (1) to get an interpretation of λ. Assume Δt is so short that in the interval $[t, t + \Delta t]$ the individual i meets only one other individual, or nobody; assume $\alpha \in (0, 1)$ is the average fraction of time we spend with other people, and $1 - \alpha$ we spend alone. If the other person is of class different from I, no infection may occur. Otherwise, there could be an infection. *Assume the probability that a person of class S gets infected, when it spends a time Δt with a person of class I, has the form $\lambda_I \Delta t$, asymptotically in $\Delta t \to 0$.* By the rule of total probability (namely we decompose the event over all possibilities depending whether subject i does not meet anybody, meets a person of class I or a person of class different from I),

$$\mathbb{P}\left(X^i\,(t + \Delta t) = I | X^i\,(t) = S\right) = \alpha \frac{I\,(t)}{N} \lambda_I \Delta t.$$

Thus

$$\lambda = \alpha \lambda_I.$$

The number λ is therefore the product between the average fraction of time we spend with other people multiplied by λ_I, the *probability per unit of time (i.e., the rate) of becoming infected during a contact with an infected person.*

The numerical value of the parameter λ_I is affected by very important factors that a little bit generically we may classify into two groups: the *contagiousness* of the virus and the *modalities of contact*. It is therefore very difficult to know λ_I in advance. One of the worst features of λ_I, and also of the other parameter α, is that they are averaged over a large variety of situations.

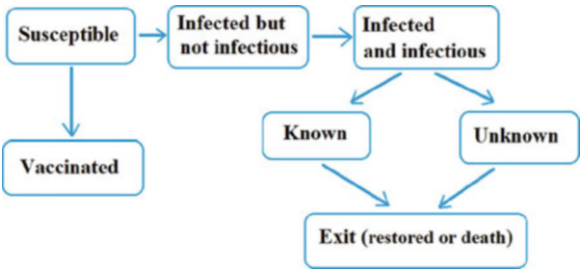
2.3 An Example of Modified SIR Model

The oversimplification of the SIR model may be, in principle, the cause of certain difficulties to give a precise meaning to parameters; or at least it is interesting to investigate more detailed compartmental models where additional phenomena can be discussed. Our particular reason to introduce one of them here is that our main multiscale model, described in Sect. 3, is an individual-based multi-class-interaction variant of this specific compartmental model.

Assume that, instead of looking only at susceptible, infected, and restored individuals, we pay attention to a more precise subdivision:

- $S(t)$ is the number of Susceptible
- $V(t)$ the number of Vaccinated
- $IN(t)$ the number of Infected but Non-infectious
- $II(t)$ the number of Infected Infectious
- $U(t)$ the number of Unknown
- $K(t)$ the number of Known
- $E(t)$ the number of Exits.

The logical diagram is



Let us explain the meaning of these groups. The first two categories are clear. The infected are divided into several categories. An infected person is non-infectious at the beginning, for a random period of some days; $IN(t)$ the number of such people. Then an infected person becomes infectious, but still does not have any symptoms and does not know to be infected; this is a very dangerous period, since the person stays in contact with other people without special care; $II(t)$ the number of such people. Then there is a bifurcation: some of them develop symptoms and are then recognized to be ill, and they are $K(t)$; others do not and remain Unknown, they are $U(t)$. It is estimated that the proportion of $U(t)$ is quite large. While the first group is treated and put in quarantine, hence in the average it is not dangerous anymore, the second group is like the group $II(t)$: very dangerous. Finally, both these categories exit the infection and enter the group $E(t)$, which essentially is the same as the restored ones of the SIR model.

One can write analogues of SIR equations for these new groups. We avoid including the vaccinated into the equations since vaccination is not a natural phenomenon, it is regulated and thus its inclusion in the differential equations is less direct. In the numerical simulations (discussed and reported below) we have taken into account the vaccinated but in a static way, namely we have considered, for instance, a population with a certain percentage of vaccinated, as an initial condition, hence excluding it from the dynamical model. The equations then have the form

$$\frac{dS(t)}{dt} = -\lambda \frac{II(t) + U(t)}{N} S(t),$$
$$\frac{dIN(t)}{dt} = \lambda \frac{II(t) + U(t)}{N} S(t) - \frac{1}{\tau_1} IN(t),$$

$$\begin{aligned}
\frac{dII(t)}{dt} &= \frac{1}{\tau_1} IN(t) - \frac{1}{\tau_2} II(t), \\
\frac{dU(t)}{dt} &= \beta \frac{1}{\tau_2} II(t) - \frac{1}{\tau_3} U(t), \\
\frac{dK(t)}{dt} &= (1-\beta) \frac{1}{\tau_2} II(t) - \frac{1}{\tau_4} K(t), \\
\frac{dE(t)}{dt} &= \frac{1}{\tau_3} U(t) + \frac{1}{\tau_4} K(t).
\end{aligned}$$

Clearly "conservation of mass" holds:

$$\frac{d(S + IN + II + U + K + E)}{dt} = 0.$$

We have many more parameters than in the SIR model. However, some of them have a clean interpretation, stability in time under different conditions, and some possibility of experimental measurement: τ_1 is the average duration of the incubation period; τ_2 the duration of the intermediate period from incubation to development of symptoms; τ_3 the average duration of infectivity for people that did not develop measurable symptoms; τ_4 the average duration of infectivity for people that developed measurable symptoms. The parameters λ and β are more difficult. The most difficult one is λ, which has a role and meaning similar to the λ of SIR. The parameter β is the proportion of cases which do not develop symptoms, a very unclear number at the beginning of the pandemic but less obscure as time goes on.

2.4 Individuals Behind the Modified SIR Model

Again we assume to have N isolated subjects $i = 1, \ldots, N$ with a "state" $X^i(t)$ now taking the following values (we do not include a dynamics of the vaccinated):

$$X^i(t) \in \{S, IN, II, U, K, E\}.$$

The probabilities of jump in very short time are (using the same short-hand notation as above)

$$\mathbb{P}\left(X^i(t+\Delta t) = IN | X^i(t) = S\right) \sim \lambda \frac{II_N(t) + U_N(t)}{N} \Delta t,$$

$$\mathbb{P}\left(X^i(t+\Delta t) = II | X^i(t) = IN\right) \sim \frac{1}{\tau_1} \Delta t,$$

$$\mathbb{P}\left(X^i\left(t+\Delta t\right)=U|X^i\left(t\right)=II\right)\sim\frac{\beta}{\tau_2}\Delta t,$$

$$\mathbb{P}\left(X^i\left(t+\Delta t\right)=U|X^i\left(t\right)=II\right)\sim\frac{1-\beta}{\tau_2}\Delta t,$$

$$\mathbb{P}\left(X^i\left(t+\Delta t\right)=E|X^i\left(t\right)=U\right)\sim\frac{1}{\tau_3}\Delta t,$$

$$\mathbb{P}\left(X^i\left(t+\Delta t\right)=E|X^i\left(t\right)=K\right)\sim\frac{1}{\tau_4}\Delta t.$$

Here we have used some of the following empirical processes:

$$S_N\left(t\right)=\sum_{j=1}^{N}1_{\{X^j(t)=S\}},\qquad IN_N\left(t\right)=\sum_{j=1}^{N}1_{\{X^j(t)=IN\}},$$

$$II_N\left(t\right)=\sum_{j=1}^{N}1_{\{X^j(t)=II\}},\qquad U_N\left(t\right)=\sum_{j=1}^{N}1_{\{X^j(t)=U\}},$$

$$K_N\left(t\right)=\sum_{j=1}^{N}1_{\{X^j(t)=K\}},\qquad E_N\left(t\right)=\sum_{j=1}^{N}1_{\{X^j(t)=E\}},$$

$$s_N\left(t\right)=N^{-1}S_N\left(t\right),\qquad in_N\left(t\right)=N^{-1}IN_N\left(t\right),$$

$$ii_N\left(t\right)=N^{-1}II_N\left(t\right),\qquad u_N\left(t\right)=N^{-1}U_N\left(t\right),$$

$$k_N\left(t\right)=N^{-1}K_N\left(t\right),\qquad e_N\left(t\right)=N^{-1}E_N\left(t\right).$$

Concerning the initial conditions we have to assume conditions analogous to those described above for SIR that we do not repeat here.

Theorem 1 *The process*

$$\left(s_N\left(t\right),in_N\left(t\right),ii_N\left(t\right),u_N\left(t\right),k_N\left(t\right),e_N\left(t\right)\right)$$

converges in probability, uniformly in time on compact intervals, to the unique solution of the equations

$$s'\left(t\right)-\lambda\left(ii\left(t\right)+u\left(t\right)\right)s\left(t\right),$$

$$in'\left(t\right)=\lambda\left(ii\left(t\right)+u\left(t\right)\right)s\left(t\right)-\frac{1}{\tau_1}in\left(t\right),$$

$$ii'(t) = \frac{1}{\tau_1} in(t) - \frac{1}{\tau_2} ii(t),$$

$$u'(t) = \frac{\beta}{\tau_2} ii(t) - \frac{1}{\tau_3} u(t),$$

$$k'(t) = \frac{1-\beta}{\tau_2} ii(t) - \frac{1}{\tau_4} k(t),$$

$$e'(t) = \frac{1}{\tau_3} u(t) + \frac{1}{\tau_4} k(t).$$

The scheme of proof is given in Appendix.

2.5 Time-Discretization

For computational purposes and maybe also for simplicity of communication, it is useful to replace the previous jump Markov processes with discrete-time Markov chains. Let us recall the idea with abstract notations. Assume we have a Markov process in continuous time $(X_t)_{t\in[0,\infty)}$ which jumps in a finite state space, time homogeneous, with transition probabilities given by

$$p_{ij}(t) = \mathbb{P}(X_t = j | X_0 = i)$$

and transition rates given by, for $j \neq i$,

$$\lambda_{ij} = \lim_{t\to 0} \frac{p_{ij}(t)}{t}.$$

Notice that probabilities take values in [0, 1] but rates are just positive, potentially larger than 1. Hence, when we write

$$p_{ij}(t) \sim \lambda_{ij} t$$

we assume that t is small enough, so that $\lambda_{ij} t \in [0, 1]$, and it is a good approximation.

Now assume we want to construct a Markov chain (discrete time) which is close to $(X_t)_{t\in[0,\infty)}$. We choose a time step $\delta > 0$ and introduce the Markov chain $(Y_n)_{n\in\mathbb{N}}$, approximation of $(X_{n\delta})_{n\in\mathbb{N}}$ with transition probabilities π_{ij} given by

$$\pi_{ij} = p_{ij}(\delta).$$

Since we do not have a precise formula for $p_{ij}(\delta)$, we approximate it by $\lambda_{ij}\delta$:

$$\pi_{ij} \sim \lambda_{ij}\delta, \tag{4}$$

taking care that we have chosen δ such that all $\lambda_{ij}\delta$ are less than 1.

In our Covid application it happens that we are interested in changing the rates λ_{ij} to explore different scenarios. In particular, to make an example, the difficult parameter λ stressed above could be doubled (roughly speaking) if certain variants take place (like the so-called delta variant). In this case, it may happen (and happens to us) that doubling λ the number $\lambda\delta$ is larger than 1. In such cases we need a more precise estimate of $p_{ij}\,(\delta)$ (in itself it is a number in [0, 1], the problem comes from its approximations).

There is, however, one case in which we have a simple exact formula for $p_{ij}\,(\delta)$, fortunately fitting precisely with our COVID example. Assume that the states i and j have only the direct communication $i \to j$, with no other multistep way to go from i to j (including loops which, from j, come back). And assume that from i we can only go to j. In the SIR example this happens for both the transitions $S \to I$ and $I \to R$. In the more elaborated model it happens for $S \to IN$, $IN \to II$, $U \to E$, $K \to E$. In this case

$$p_{ij}\,(\delta) = 1 - e^{-\lambda_{ij}\delta}$$

(of which $\lambda_{ij}\delta$ is the first order approximation for small δ). In this formula we can double the rate λ_{ij} without any trouble.

As a numerical example: assume that $\pi_{ij} = 0.65$. It means $1 - e^{-\lambda_{ij}\delta} = 0.65$, namely $\lambda_{ij}\delta = 1.05$; doubling λ_{ij} we get $2\lambda_{ij}\delta = 2.1$, hence $1 - e^{-2\lambda_{ij}\delta} = 0.877$, namely $\pi_{ij}^{\text{new}} = 0.877$.

There is a second element we want to clarify about the connection between continuous and discrete systems: the links between the transition parameters and the waiting times. Consider a state i, for a jump Markov processes in continuous time $(X_t)_{t\in[0,\infty)}$ and call T_i^{cont} the first time the process jumps from i to another state; it is exponentially distributed. We have

$$\sum_{j\neq i} \lambda_{ij} = \frac{1}{\mathbb{E}\left[T_i^{\text{cont}}\right]}.$$

Consider now the case of Markov chain $(Y_n)_{n\in\mathbb{N}}$ and call T_i^{disc} the first discrete step (greater or equal than one) the process jumps from i to another state; it is geometrically distributed. Notice that $\mathbb{E}\left[T_i^{\text{disc}}\right] \geq 1$, since $T_i^{\text{disc}} \geq 1$, property not shared by T_i^{cont}. Then

$$\sum_{j\neq i} \pi_{ij} = \frac{1}{\mathbb{E}\left[T_i^{\text{disc}}\right]}.$$

These rules are fundamental to fit parameters, since it is easier to have information on waiting times. But which is the link between the continuous and the discrete rule?

Assume the time step of discretization is $\delta > 0$. Then, roughly speaking,

$$T_i^{\text{cont}} \sim \delta T_i^{\text{disc}},$$

hence $\mathbb{E}\left[T_i^{\text{cont}}\right] \sim \delta \mathbb{E}\left[T_i^{\text{disc}}\right]$, hence

$$\sum_{j \neq i} \pi_{ij} \sim \delta \sum_{j \neq i} \lambda_{ij},$$

which is coherent with the rule (4) decided above. In the next section we describe a model taken from [15], in discrete time, where all transition probabilities have been fitted by using their relations with expected waiting times like T_i^{disc}; here revisiting [15], we shall use (4). The discussion just clarifies that there is no mistake: fitting π_{ij} by means of inverse average waiting times or by means of (4) is equivalent.

3 The Individual-Based Model of [15]

This section is devoted to a Markov chain individual-based model developed in [15]. It is an elaboration of the model of Section 2.4, but, in order to clarify its genesis, let us say that the model of Sect. 2.4 appears in this chapter for the first time, with the purpose to create a closer link with SIR; obviously, on the other hand, SIR and generically its variants have been a fundamental source of inspiration in the development of [15].

First, [15] is a discrete-time analog of the model of Sect. 2.4, with transition probabilities chosen according to the rules of Sect. 2.5. However, the model of [15] makes the second step forward from the viewpoint of realism: at the individual level, [15] introduces *different interaction structures*. Let us explain this key issue.

In both the simple individual-based models of Sects. 2.2 and 2.4, the interactions are mean field, global: each individual interacts with each other. This is essential for having the simple mean field macroscopic description based on the systems of ordinary differential equations of Sects. 2.1 and 2.3; and the proof of convergence outlined in the Appendix is so simple because of the mean field structure. Said in another way, the probability of getting infected is proportional to the global number of infected. This is not realistic: people spend a large amount of time inside home, in contact only with members of our family; then they spend another amount of time in contact with people of a very restricted group (work or study colleagues, friends; always the same people); finally, they spend a usually small amount of time in contact with generic people (during public transportations, at restaurants, shops, museums, etc.). It is unrealistic that the totality of infected people affects a single person of class S. The work [15] thus introduces a decomposition according to these three classes.

Concerning the problem of knowledge of the parameter λ, let us remark that it is this decomposition only, and not the fact that we deal with the modified SIR,

which gives us some opportunity to enter the meaning of λ and replace it by more elementary parameters.

Let us describe in more detail the interaction structures used in [15]. One class is the family. It is very different from everything else because, in the average over the population, one spends much more time in the family and the level of protections (masks, distancing) is almost absent there (however, after the emergence of COVD-19 most people clean hands and objects much more than before, with positive effects also on the family). The family is formalized by specifying, for each individual $i = 1, \ldots, N$, a set of individuals

$$J_{\text{fam}}(i) \subset \{1, \ldots, N\}$$

which does not contain i itself. These data are collected into an $N \times N$ matrix A^{fam}, with 0 and 1 entries: for every $i = 1, \ldots, N$, we set $A_{ij}^{\text{fam}} = 1$ if $j \in J_{\text{fam}}(i)$, zero otherwise.

The second class is the city in its complex. When we enter shops, supermarkets, when we take public transportations, go to a stadium or a church or a restaurant and so on, we get in contact with people that usually we do not know, generic people of the city where we live. These contacts are very different from those of the family; for instance, under special rules imposed by a government, the time spent in these contacts can be very reduced; and the modalities are very affected by protection rules (masks, distancing, cleanings). We do not need to formalize these contacts with a matrix, but for analogy with the previous case, we introduce an $N \times N$ matrix A^{city}, with all entries equal to 1.

The third class is the group of friends and colleagues which we usually meet, excluding the family members already considered above. Again certain restrictions apply to this third group, not necessarily equal to the restrictions of the city. But the main reason to introduce this third class is not the difference in parameters with respect to the city (which can be small), but the fact that spending considerable time in such groups with respect to the full city community leads to a potential limitation in the spread of the virus, a fact that we wanted to investigate. In [15] these groups were called pseudo-bubbles, perhaps because for a period in 2020 it was common to refer to bubbles describing a governmental rule in Belgium which allowed people to meet only a certain number of friends. Here we shall simply use the name "groups." The groups are formalized by specifying, for each individual $i = 1, \ldots, N$, a set of individuals

$$J_{\text{group}}(i) \subset \{1, \ldots, N\}$$

which does not contain i itself and collected them into an $N \times N$ matrix A^{group} such that, for every $i = 1, \ldots, N$, we set $A_{ij}^{\text{group}} = 1$ if $j \in J_{\text{group}}(i)$, zero otherwise.

Therefore the three different *networks* of contacts are described by the matrices A^{fam}, A^{group}, A^{city}. In addition we introduce three parameters $\alpha_{\text{fam}}, \alpha_{\text{group}}, \alpha_{\text{city}} \in (0, 1)$ which represent the fractions of time a generic individual meets, respectively,

people of the family, of the group and of the city; *they do not sum to one*, because there is a fraction of time an individual does not meet anybody.

By the rule of total probability, now we have (compare with (1) above)

$$\begin{aligned}
&\mathbb{P}\left(X^i(t+\Delta t) = I | X^i(t) = S\right) \qquad (5)\\
&\sim \alpha_{\text{fam}} \lambda_{I,\text{fam}} \frac{I^i_{\text{fam}}(t)}{N^i_{\text{fam}}} \Delta t\\
&+ \alpha_{\text{group}} \lambda_{I,\text{group}} \frac{I^i_{\text{group}}(t)}{N^i_{\text{group}}} \Delta t\\
&+ \alpha_{\text{city}} \lambda_{I,\text{city}} \frac{I^i_{\text{city}}(t)}{N^i_{\text{city}}} \Delta t.
\end{aligned}$$

The parameters $\lambda_{I,\text{fam}}$, $\lambda_{I,\text{group}}$, $\lambda_{I,\text{city}}$ are the *probabilities per unit of time of getting the infection during a contact with an infected person, respectively, of the family, of the personal group and of the city*.

The random quantities $I^i_{\text{fam}}(t)$, $I^i_{\text{group}}(t)$, $I^i_{\text{city}}(t)$ and the numbers N^i_{fam}, N^i_{group}, N^i_{city} are defined as

$$\begin{aligned}
I^i_{\text{fam}}(t) &= \sum_{j: A^{\text{fam}}_{ij}=1} 1_{\{X^j(t)=I\}},\\
N^i_{\text{fam}} &= \sum_{j: A^{\text{fam}}_{ij}=1} 1,
\end{aligned}$$

and similarly for the other quantities. In fact, in the simulations of [15] and also the new ones reported here, the numbers N_{fam}, N_{group}, N_{city} are chosen independent of i, for simplicity.

The potential advantage of this subdivision of the contacts is that we have a more precise interpretation of the parameters α_{fam}, α_{group}, α_{city}, $\lambda_{I,\text{fam}}$, $\lambda_{I,\text{group}}$, $\lambda_{I,\text{city}}$, compared to α , λ_I. Let us make an example. At the beginning of the epidemic people were not aware of it and did not take any protection measures. Assume it is possible to extract a fit of the parameter λ_I in such a situation (unfortunately also the data about the infection in that period are corrupted by the surprise). Then this value is quite representative of $\lambda_{I,\text{fam}}$, the case where still now we do not apply protection rules. Another example refers to the time parameters. Guessing a priori the global time α of the SIR model is very vague. But guessing a priori the times associated with different classes looks easier. Less ambitiously, in the new model it is easier to investigate at the qualitative level the effect of reduction of the time spent in certain classes due to rules imposed by governments: we know which specific parameter we have to affect, not generically α in a measure that is not clear.

3.1 A Formula for the Parameter λ of Compartmental Models

Assume the two models, given by the two formulae (1) and (5), are both relatively realistic. Then, with some degree of approximation,

$$\begin{aligned}\lambda\frac{I\,(t)}{N} \sim\ & \alpha_{\text{fam}}\lambda_{I,\text{fam}}\frac{I_{\text{fam}}\,(t)}{N_{\text{fam}}}\\ &+\alpha_{\text{group}}\lambda_{I,\text{group}}\frac{I_{\text{group}}\,(t)}{N_{\text{group}}}\\ &+\alpha_{\text{city}}\lambda_{I,\text{city}}\frac{I_{\text{city}}\,(t)}{N_{\text{city}}},\end{aligned}$$

where now the symbol $\sim$ refers to the approximation, not to a limit as $\Delta t \to 0$. Therefore

$$\begin{aligned}\lambda \sim\ & \alpha_{\text{fam}}\lambda_{I,\text{fam}}\frac{I_{\text{fam}}\,(t)}{I\,(t)}\frac{N}{N_{\text{fam}}}\\ &+\alpha_{\text{group}}\lambda_{I,\text{group}}\frac{I_{\text{group}}\,(t)}{I\,(t)}\frac{N}{N_{\text{group}}}\\ &+\alpha_{\text{city}}\lambda_{I,\text{city}}\frac{I_{\text{city}}\,(t)}{I\,(t)}\frac{N}{N_{\text{city}}}.\end{aligned}$$

Assume, for simplicity, that the infected people in the average are divided into three classes uniformly, namely

$$\begin{aligned}\frac{I_{\text{fam}}\,(t)}{I\,(t)} &\sim \frac{N_{\text{fam}}}{N},\\ \frac{I_{\text{group}}\,(t)}{I\,(t)} &\sim \frac{N_{\text{group}}}{N},\\ \frac{I_{\text{city}}\,(t)}{I\,(t)} &\sim \frac{N_{\text{city}}}{N},\end{aligned}$$

(if these were exact equalities, two of them imply the third one). Then:

Claim 2 *Approximately, we have*

$$\lambda \sim \alpha_{fam}\lambda_{I,fam} + \alpha_{group}\lambda_{I,group} + \alpha_{city}\lambda_{I,city}. \tag{6}$$

This is a simple rule which allows one to use the SIR model as a byproduct of the individual-based analysis. The approximate equation (6) is a sort of total probability formula for transition rates.

3.2 Analysis of the Fluctuations

In this section we want to understand the fluctuations of the number of infected people. By experience, in this COVID-19 epidemic, we have seen that this number is small compared to the global population (although very high in terms of human cost and public health system capability). The fluctuations of a small quantity (small compared to the cardinality of the full system) is a very delicate issue.

Let us start with an elementary model of fluctuations: assume we have independent identically distributed random variables $X_1, \dots, X_N$, one for each subject of the population, such that $X_i = 1$ if individual i is infected, $X_i = 0$ if not. r.v.'s X_i are Bernoulli with parameter $i\,(t)$, in the notations above: $i\,(t)$ is the percentage of infected people, hence interpreted as the probability to be infected. The number of infected

$$S_N := X_1 + \ldots + X_N$$

is a Binomial r.v. with parameters N and $i\,(t)$. Consider for instance the case

$$N = 100000,$$

$$i\,(t) = \frac{500}{100000} = 0.005,$$

which corresponds to several Italian cities during the Spring 2020 epidemic. The histogram of ten thousand random values from such a distribution are plotted in Fig. 1. The fluctuations, as we see, are quite small, of the order of a few dozens. They can be guessed by the Central Limit Theorem (CLT):

$$S_N - Ni\,(t) \sim \sqrt{i\,(t)\,(1 - i\,(t))}\sqrt{N},$$

where $i\,(t) = \frac{500}{N}$, hence

$$S_N - Ni\,(t) \sim \sqrt{500} \sim 22.$$

This result is completely unrealistic compared to real data, reported in Sect. 3.3. The well-known "concentration" property of Binomial distributions (related to the CLT) is unrealistic in this real case. Where does the additional spread come from?

Theoretically, the mean field limit, with the consequent propagation of chaos property, is correct for the simplified individual-based model described in Sect. 2.1. What is wrong, then? A secondary reason could be the fact that the real situation is not of mean field type: already in our modified model of Sect. 2.4 we have two non-mean-field levels of interactions, the family and the personal group, which a priori could break the propagation of chaos. However, it is difficult to believe that the true reason is so sophisticated. The main reason in our opinion is that *the smallness of the infected group breaks the mean field property*. Let us explain this important concept in some detail.

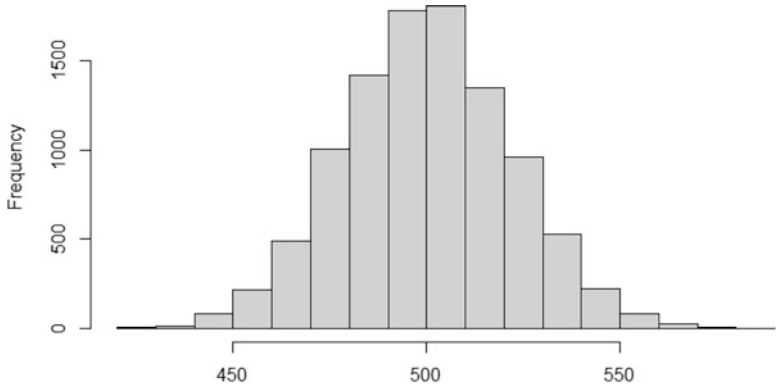


Fig. 1 Ten thousand random values from Binomial distribution

Recall the specification of the initial conditions (3). It requires that the number of infected individuals at time zero (and therefore also later on) is a non-negligible *fraction* of N. The concrete number 500/100000 above is not a fraction of N, in practical terms (10,000 is a fraction of N). If $I(t)$ is a fraction β of N, then $i(t) = \beta$ and

$$S_N - Ni(t) \sim \sqrt{\beta(1-\beta)}\sqrt{N}.$$

Assume, for instance, $\sqrt{\beta(1-\beta)} = \sqrt{\frac{1}{10}\left(1-\frac{1}{10}\right)} = 0.3$. Then

$$S_N - Ni(t) \sim 0.3 \cdot \sqrt{100000} \sim 95,$$

which is a very realistic estimate of the real data!

But in our case β is not $\frac{1}{10}$, it is 0.005. Hence we are not under the assumptions of the mean field result. How to estimate the uncertainty if we cannot apply the paradigms of the mean field theory? The easiest is a direct numerical simulation of the individual-based model. We have done it in our modified model and the results are astonishingly close to reality, as shown in the next section.

3.3 Simulations

In this section we present some of the simulations from [15] that we deem to be the most relevant. The first one is the distribution of the incubation period. We choose this one because we think it is fundamental that a model can simulate correctly some significant experimental quantities, such as incubation period and serial interval, and their distributions, besides the number of cases. These quantities characterize

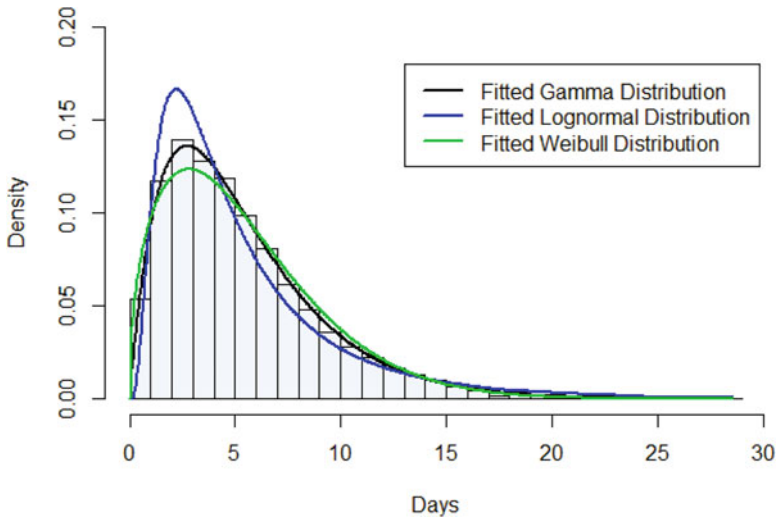


Fig. 2 Histogram obtained calling the function that calculates the incubation period 5000 times

the illness and its infectiousness and are therefore truly important to consider when studying the contagion's dynamics. In order to obtain Fig. 2, we used a Monte Carlo method to extract a large sample of incubation periods and we used the results to make an histogram and to fit different distributions. What we obtain is incredibly close to results we can find in literature, see [19], for example, estimated from real cases.

Another result we obtained that we think is worth being presented here, as we were saying in the previous section, is the simulation of the total number of cases from March 2020 (beginning of the Pandemics) to May 2020 (when some of the restrictions were abolished). This period of time is really characteristic because it contains the first phase when there wasn't any restrictions or any need to wear personal protective equipment and the second "lockdown" phase. To differentiate these two phases in our simulations, we decided to change the parameters of the model, specifically the ones taking into account the time spent in contact with individuals other than those in the family, and the ones regulating the probability of getting infected when in contact with an infectious individual (this probability is proven to be reduced by face masks and social distancing).

The first figure represents the total number of diagnosed cases of every of the ten Tuscany's provinces, scaled to an appropriate number of individuals and with a black line representing the average. These real data are taken from the page of Franco Mossotto.[1]

[1] https://datastudio.google.com/u/0/reporting/91350339-2c97-49b5-92b8-965996530f00/page/RdlHB.

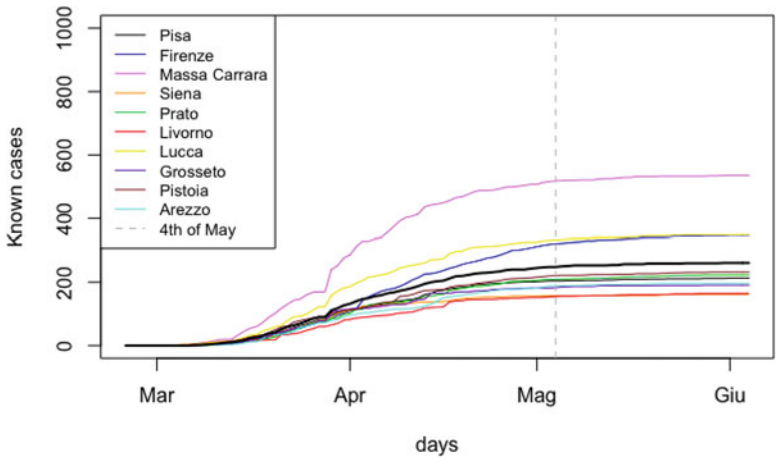


Fig. 3 Total cases of Tuscany's provinces

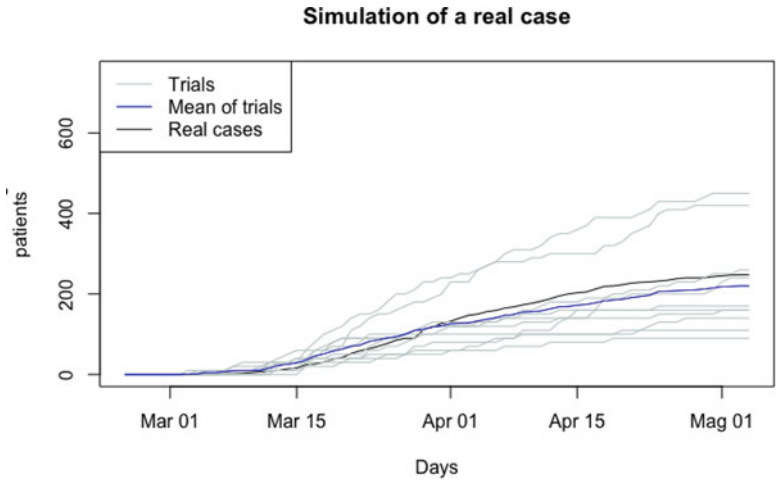


Fig. 4 Fit of mean using only families

The reason we chose Tuscany to have a benchmark for our simulations is because there the number of cases was like an enormous controlled experiment. We can observe from Fig. 3 that the variability of these real cases goes about from 100 to 500. Hence we expect that a good fit of the mean is given by trials with variability in this range. Notice that the majority of the lines stay under the mean. In Fig. 4 we try to fit the mean using only the families as social barriers.

We supposed that during the pre-lockdown period the parameter λ that we fitted as explained in [15], was the same for all the situations, as this value corresponds to the biological one. During the lockdown phase it is reduced to $\lambda/4$ in the city, in order to take into account the constraints imposed by the government about face

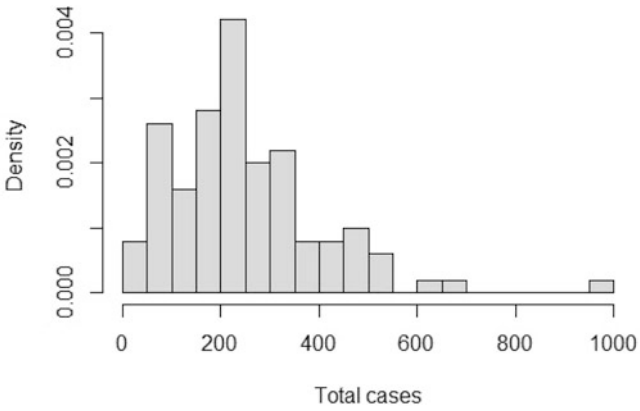


Fig. 5 Total cases on the 4th of May in 100 trials. Model without groups

masks and social distancing. Similarly, the average proportion of time spent in the city or, equivalently, in contact with individuals outside the family, is reduced when the lockdown begins, going from $\alpha_{\text{city}} = 0.5$ to $\alpha_{\text{city},1} = 0.025$ for a part of the population (that we imagined as those people that did not go outside their house for reasons other than grocery shopping) and to $\alpha_{\text{city},1} = 0.35$ for the remaining part of the population (people that still needed to go to work). It is noteworthy that the variability is a little bit wider if compared to the real cases, especially for the lower extreme of the range. However, it is quite close to what we wanted to obtain and it is also possible to see the "long tail" that characterizes the real cases. This becomes even more evident in Fig. 5, where we have plotted an histogram of the total cases in the last day, that is the 4th of May, making one-hundred trials of our code.

Using the uniform scenario it is not easy to fit the mean because the introduction of the families slows down a lot the epidemics. Differently, introducing the groups we managed to fit the mean, introducing more realism. However, more realism goes together with more parameters, thus for the period under exam is sufficient to use the second scenario to model the real cases.

3.4 Presence of Immunized Population and Virus Variants

In this section we present the results we obtained while trying to answer one of the most difficult questions: what would happen if in a situation where a certain proportion of the population is vaccinated and the virus is no longer circulating, a new infected (and infectious) individual appears?

In order to make an attempt we consider different percentage of immunized individuals, where immunized refers to individuals who recovered from Covid-19 or people that have been vaccinated. A small portion of vaccinated individuals can

still be infected, but their infectiousness is lower and we made the simplification that they have nearly no active role in infection spreading.

We try to simulate this situation using the model described in [15]. We choose two different percentages of immunized population, 30 and 70%, to have an idea of what happens when a small proportion of the population is immunized and how it changes when we have much more immune individuals. Further, for each percentage of immunized population, we ran two simulations of our model with different values of λ. We want to show that, in addition to the immunized population, social distancing and masks lead to really different scenarios. Basically, those scenarios can be described with the results of one-hundred trial of our code, where we set a random individual to be infected. The results show that, with containing measures

$$\mathbb{P}(T > 100) = 0.01,$$

in Fig. 8, whereas in the scenario of Fig. 9

$$\mathbb{P}(T > 100) = 0.2.$$

On the other hand, in a situation of absence of containing measures, Fig. 6 shows that, despite an high number of immunized population,

$$\mathbb{P}(T > 100) = 0.24$$

and with a 30% of vaccinated individuals (Fig. 7)

$$\mathbb{P}(T > 100) = 0.79,$$

where T are the total cases registered after 120 days.

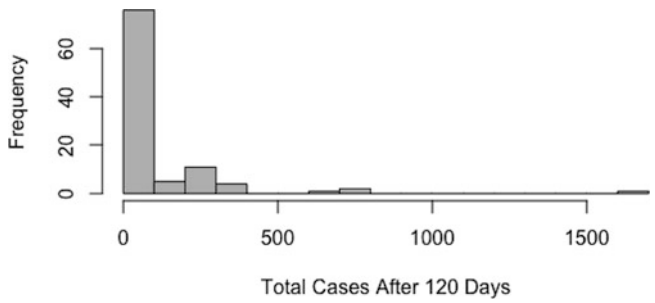


Fig. 6 Scenario with $N = 100{,}000$ individuals and $G = 120$ days. No social distancing or masks and 70% of immunized population

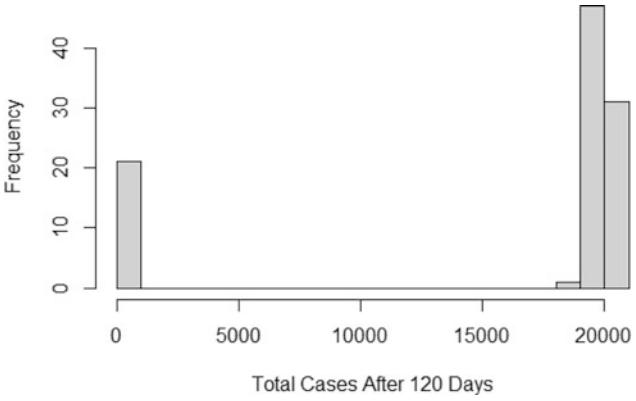


Fig. 7 Scenario with $N = 100{,}000$ individuals and $G = 120$ days. No social distancing or masks and 30% of immunized population

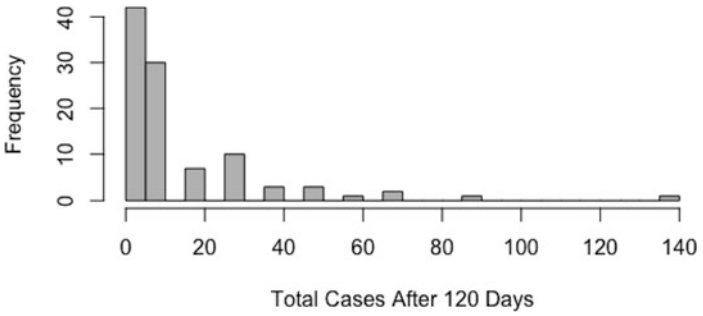


Fig. 8 Scenario with $N = 100{,}000$ individuals and $G = 120$ days. Social distancing and masks and 70% of immunized population

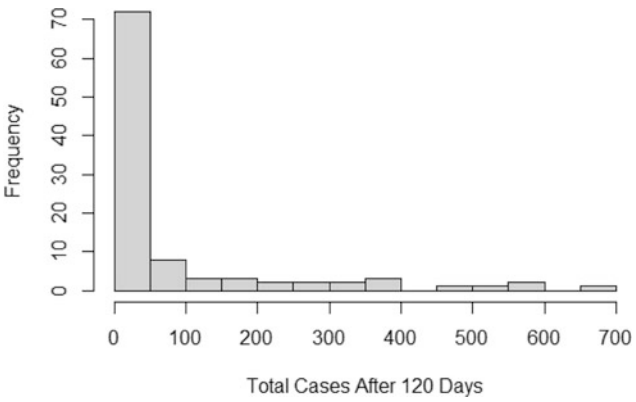


Fig. 9 Scenario with $N = 100{,}000$ individuals and $G = 120$ days. Social distancing and masks and 30% of immunized population

Appendix

In this appendix we describe a few rigorous details about the link between compartmental models and their individual-based counterparts. We do this for the modified model of Sect. 2.3, which is new, so the result of the present section is original.

The best way to deal with continuous time jump Markov processes and their macroscopic limit is, following [23], by introducing the state space

$$\Xi_N = \{S, IN, II, U, K, E\}^N$$

and a Markov process $X_t \in \Xi_N$, $X_t = \left(X_t^1, \ldots, X_t^N\right)$, the empirical quantities

$$S_N(t) = \sum_{j=1}^{N} 1_{\{X^j(t)=S\}}, \qquad IN_N(t) = \sum_{j=1}^{N} 1_{\{X^j(t)=IN\}},$$

$$II_N(t) = \sum_{j=1}^{N} 1_{\{X^j(t)=II\}}, \qquad U_N(t) = \sum_{j=1}^{N} 1_{\{X^j(t)=U\}},$$

$$K_N(t) = \sum_{j=1}^{N} 1_{\{X^j(t)=K\}}, \qquad E_N(t) = \sum_{j=1}^{N} 1_{\{X^j(t)=E\}},$$

$$s_N(t) = N^{-1} S_N(t), \qquad in_N(t) = N^{-1} IN_N(t),$$

$$ii_N(t) = N^{-1} II_N(t), \qquad u_N(t) = N^{-1} U_N(t),$$

$$k_N(t) = N^{-1} K_N(t), \qquad e_N(t) = N^{-1} E_N(t),$$

and the infinitesimal generator $\mathcal{L}$, defined on test functions $F : \Xi_N \to \mathbb{R}$, as

$$\begin{aligned}(\mathcal{L}F)(X) = &\left(\mathcal{L}_{S,IN}F\right)(X) + \left(\mathcal{L}_{IN,II}F\right)(X) + \left(\mathcal{L}_{II,U}F\right)(X) + \left(\mathcal{L}_{II,K}F\right)(X) \\ &+ \left(\mathcal{L}_{U,E}F\right)(X) + \left(\mathcal{L}_{K,E}F\right)(X),\end{aligned}$$

where

$$\left(\mathcal{L}_{S,IN}F\right)(X) = \sum_{i=1,\ldots,N} 1_{\{X^i=S\}} \lambda \left(ii_N + u_N\right) \left(F\left(X^{i,S\to IN}\right) - F(X)\right),$$

$$\left(\mathcal{L}_{IN,II}F\right)(X) = \sum_{i=1,\ldots,N} 1_{\{X^i=IN\}} \frac{1}{\tau_1} \left(F\left(X^{i,IN\to II}\right) - F(X)\right),$$

$$\left(\mathcal{L}_{II,U}F\right)(X) = \sum_{i=1,\dots,N} 1_{\{X^i=II\}} \frac{\beta}{\tau_2}\left(F\left(X^{i,II\to U}\right) - F(X)\right),$$

$$\left(\mathcal{L}_{II,K}F\right)(X) = \sum_{i=1,\dots,N} 1_{\{X^i=II\}} \frac{1-\beta}{\tau_2}\left(F\left(X^{i,II\to K}\right) - F(X)\right),$$

$$\left(\mathcal{L}_{U,E}F\right)(X) = \sum_{i=1,\dots,N} 1_{\{X^i=U\}} \frac{1}{\tau_3}\left(F\left(X^{i,U\to E}\right) - F(X)\right),$$

$$\left(\mathcal{L}_{K,E}F\right)(X) = \sum_{i=1,\dots,N} 1_{\{X^i=K\}} \frac{1}{\tau_4}\left(F\left(X^{i,K\to E}\right) - F(X)\right).$$

Here $X \in \Xi_N$, $X = \left(X^1, \dots, X^N\right)$, $X^{i,S\to IN}$ is configuration X where, if X^i is different from S, nothing changes, namely $X^{i,S\to IN} = X$ (but this case is inessential because there in the multiplication by $1_{\{X^i=S\}}$). But if $X^i = S$, then we replace S by IN in position X^i, namely $\left(X^{i,S\to IN}\right)^i = IN$, $\left(X^{i,S\to IN}\right)^j = X^j$ for all $j \neq i$. The definitions of $X^{i,IN\to II}$, etc. are similar. Moreover, we have used the notations ii_N and u_N, defined similarly to $ii_N(t)$ and $u_N(t)$, but referring only to the single configuration X instead of the stochastic process $X(t)$.

With these preliminaries, one can apply Dynkin formula ([23, Appendix 1, Section 5]) and get

$$s_N(t) = s_N(0) - \lambda \int_0^t (ii_N(s) + u_N(s))\, s_N(s)\, ds + M_N^s(t),$$

$$in_N(t) = in_N(0) + \int_0^t \left(\lambda (ii_N(s) + u_N(s))\, s_N(s) - \frac{1}{\tau_1} in_N(s)\right) ds + M_N^{in}(t),$$

$$ii_N(t) = ii_N(0) + \int_0^t \left(\frac{1}{\tau_1} in_N(s) - \frac{1}{\tau_2} ii_N(s)\right) ds + M_N^{ii}(t),$$

$$u_N(t) = u_N(0) + \int_0^t \left(\frac{\beta}{\tau_2} ii_N(s) - \frac{1}{\tau_3} u_N(s)\right) ds + M_N^{u}(t),$$

$$k_N(t) = k_N(0) + \int_0^t \left(\frac{1-\beta}{\tau_2} ii_N(s) - \frac{1}{\tau_4} k_N(s)\right) ds + M_N^{k}(t),$$

$$e_N(t) = e_N(0) + \int_0^t \left(\frac{1}{\tau_3} u_N(s) + \frac{1}{\tau_4} k_N(s)\right) ds + M_N^{e}(t),$$

where $M_N^s(t)$, $M_N^{in}(t)$, $M_N^{ii}(t)$, $M_N^u(t)$, $M_N^k(t)$, $M_N^e(t)$ are martingales, the quadratic variation of which is given by a further application of Dynkin formula,

that we omit (being the final result about the estimates from above essentially equal in most cases, see the second formula of Lemma 5.1 in [23, Appendix 1]), which lead to

$$\mathbb{E}\left[\left|M_N^s(t)\right|^2\right] \leq \frac{C}{N},$$

and similarly for the other martingales. Let us show where the first identity above comes from; the others are similar.

Consider the functional $F_{j,S} : \Xi_N \to \mathbb{R}$ given by $F_{j,S}(X) = 1_{\{X^j=S\}}$, where $X \in \Xi_N$, $X = \left(X^1, \ldots, X^N\right)$. Let us compute

$$\begin{aligned}\left(\mathcal{L}_{S,IN}F_{j,S}\right)(X) &= \sum_{i=1,\ldots,N} 1_{\{X^i=S\}}\lambda\left(ii_N + u_N\right)\left(F_{j,S}\left(X^{i,S\to IN}\right) - F_{j,S}(X)\right)\\ &= \sum_{i=1,\ldots,N} 1_{\{X^i=S\}}\lambda\left(ii_N + u_N\right)\left(1_{\left\{\left(X^{i,S\to IN}\right)^j=S\right\}} - 1_{\{X^j=S\}}\right).\end{aligned}$$

In the sum, consider the addend i. If $i \neq j$, in the bracket we have

$$1_{\left\{\left(X^{i,S\to IN}\right)^j=S\right\}} - 1_{\{X^j=S\}} = 1_{\{X^j=S\}} - 1_{\{X^j=S\}} = 0,$$

recalling the rule $\left(X^{i,S\to IN}\right)^j = X^j$. Therefore

$$\left(\mathcal{L}_{S,IN}F_{j,S}\right)(X) = 1_{\{X^j=S\}}\lambda\left(ii_N + u_N\right)\left(1_{\left\{\left(X^{j,S\to IN}\right)^j=S\right\}} - 1_{\{X^j=S\}}\right).$$

Moreover, $\left(X^{j,S\to IN}\right)^j = IN$ by the rule above, and $1_{\{IN=S\}} = 0$, hence

$$\left(\mathcal{L}_{S,IN}F_{j,S}\right)(X) = -\lambda\left(ii_N + u_N\right)1_{\{X^j=S\}}.$$

Now, calling $s_N = s_N(X)$ the functional $s_N = N^{-1}\sum_{j=1}^N F_{j,S}$, we have

$$\left(\mathcal{L}_{S,IN}s_N\right)(X) = -\lambda\left(ii_N + u_N\right)s_N.$$

By similar arguments we get

$$\left(\mathcal{L}_{IN,II}s_N\right)(X) = 0$$

and so on for all the others. Indeed, as above,

$$\left(\mathcal{L}_{IN,II}F_{j,S}\right)(X) = \sum_{i=1,\ldots,N} 1_{\{X^i=IN\}}\frac{1}{\tau_1}\left(1_{\left\{\left(X^{i,IN\to II}\right)^j=S\right\}} - 1_{\{X^j=S\}}\right)$$

$$= 1_{\{X^i = IN\}} \frac{1}{\tau_1} \left(1_{\left\{ \left(X^{j, IN \to II} \right)^j = S \right\}} - 1_{\{X^j = S\}} \right)$$

$$= -\frac{1}{\tau_1} 1_{\{X^i = IN\}} 1_{\{X^j = S\}} = 0.$$

Summarizing,

$$(\mathcal{L} s_N)(X) = -\lambda (ii_N + u_N) s_N.$$

Now, Dynkin formula tells us that

$$F(X(t)) = F(X(0)) + \int_0^t (\mathcal{L}F)(X(s))\, ds + M(t),$$

where $M(t)$ is a martingale. Using the functional $F = s_N$ we get

$$s_N(t) = s_N(0) - \lambda \int_0^t (ii_N(s) + u_N(s)) s_N(s)\, ds + M_N^s(t).$$

The proof of the other equations is similar.

Consider now the process

$$(s_N(t), in_N(t), ii_N(t), u_N(t), k_N(t), e_N(t))$$

on the Skorohod space on a generic time interval $[0, T]$. Using a classical tightness criterium of Aldous [23, Chapter 4, Proposition 1.6], one can prove that the laws P_N of these 6-dimensional processes form a tight family on Skorohod space, also with respect to the uniform topology. Thus there exists a weakly convergent subsequence $\left(P_{N_k}\right)_{k \in \mathbb{N}}$, in the uniform topology; call P its limit. One can show it is supported on continuous functions. By an easy argument of weak convergence based on the functionals [23, Chapter 4, Section 2, Second step],

$$\Phi_S \left(1 \wedge \sup_{t \in [0,T]} \left| s(t) - s(0) + \lambda \int_0^t (ii(s) + u(s)) s(s)\, ds \right| \right),$$

$$\Phi_{IN} \left(1 \wedge \sup_{t \in [0,T]} \left| s(t) - s(0) + \int_0^t \left(\lambda (ii(s) + u(s)) s(s) - \frac{1}{\tau_1} in(s) \right) ds \right| \right),$$

and Φ_{II}, Φ_U, Φ_K, Φ_E similarly defined, one proves that P is supported on continuous solutions of the system (mean square convergence to zero of the martingales is used here)

$$s(t) = s(0) - \lambda \int_0^t (ii(s) + u(s))\, s(s)\, ds,$$

$$in(t) = in(0) + \int_0^t \left(\lambda (ii(s) + u(s))\, s(s) - \frac{1}{\tau_1} in(s)\right) ds,$$

$$ii(t) = ii(0) + \int_0^t \left(\frac{1}{\tau_1} in(s) - \frac{1}{\tau_2} ii(s)\right) ds,$$

$$u(t) = u(0) + \int_0^t \left(\frac{\beta}{\tau_2} ii(s) - \frac{1}{\tau_3} u(s)\right) ds,$$

$$k(t) = k(0) + \int_0^t \left(\frac{1-\beta}{\tau_2} ii(s) - \frac{1}{\tau_4} k(s)\right) ds,$$

$$e(t) = e(0) + \int_0^t \left(\frac{1}{\tau_3} u(s) + \frac{1}{\tau_4} k(s)\right) ds.$$

By classical Cauchy theorem, this system has a unique solution, hence P is supported on its single solution, namely P is a delta Dirac measure. Therefore the whole sequence P_N converges in law to P, and in fact converges in probability.

References

1. N.T. Bailey, et al., *The Mathematical Theory of Infectious Diseases and its Applications* (Charles Griffin, Glasgow, 1975)
2. N. Bellomo, L. Gibelli, N. Outada, On the interplay between behavioral dynamics and social interactions in human crowds. Kinetic Rel. Models **12**, 397–409 (2019)
3. N. Bellomo, K.J. Painter, Y. Tao, M. Winkler, Occurrence vs. absence of taxis-driven instabilities in a May–Nowak model for virus infection. SIAM J. Appl. Math. **79**(5), 1990–2010 (2019)
4. N. Bellomo, R. Bingham, M.A. Chaplain, G. Dosi, G. Forni, D.A. Knopoff, J. Lowengrub, R. Twarock, M.E. Virgillito, A multi-scale model of virus pandemic: heterogeneous interactive entities in a globally connected world. Math. Models Methods Appl. Sci. **30**, 1591–1651 (2020)
5. V. Capasso, L. Maddalena, Convergence to equilibrium states for a reaction-diffusion system modelling the spatial spread of a class of bacterial and viral diseases. J. Math. Biol. **13**(2), 173–184 (1981)
6. V. Capasso, S. Paveri-Fontana, A mathematical model for the 1973 cholera epidemic in the European Mediterranean region. Revue d'épidémiologie et de Santé Publiqué **27**(2), 121–132 (1979)
7. V. Capasso, E. Grosso, G. Serio, Mathematical models in epidemiological studies. I. Application to the epidemic of cholera verified in Bari in 1973. Annali Sclavo; Rivista di Microbiologia e di Immunologia **19**, 193–208 (1977)
8. M. Cecconi, G. Forni, A. Mantovani, Covid-19: summer 2020 report. Accademia Nazionale dei Lincei, Commissione Salute (2020)

9. M. Cecconi, G. Forni, A. Mantovani, Ten things we learned about covid-19. Intensive Care Med. **46**(8), 1590–1593 (2020)
10. B.R. Craig, T. Phelan, J.P. Siedlarek, J. Steinberg, Improving epidemic modeling with networks. Federal Reserve Bank of Cleveland, Econ. Commentary 2020–23 (2020) https://doi.org/10.26509/frbc-ec-202023
11. D.J. Daley, J. Gani, *Epidemic Modelling: An Introduction*, vol. 15 (Cambridge University Press, Oxford, 2001)
12. M. Day, Covid-19: four fifths of cases are asymptomatic, China figures indicate. BMJ. 2020 Apr 2;369:m1375. https://doi.org/10.1136/bmj.m1375 PMID: 32241884
13. O. Diekmann, Limiting behaviour in an epidemic model. Nonlinear Anal. Theory Methods Appl. **1**(5), 459–470 (1977)
14. O. Diekmann, H. Heesterbeek, T. Britton, *Mathematical Tools for Understanding Infectious Disease Dynamics*, vol. 7 (Princeton University Press, Princeton, 2012)
15. F. Flandoli, E. LaFauci, M. Riva, Individual-based Markov model of virus diffusion: comparison with covid-19 incubation period, serial interval and regional time series. Math. Models Methods Appl. Sci.**31**(5), 907–939 (2021)
16. S. Flaxman, S. Mishra, A. Gandy, H.J.T. Unwin, T.A. Mellan, H. Coupland, C. Whittaker, H. Zhu, T. Berah, J.W. Eaton, et al., Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. Nature **584**(7820), 257–261 (2020)
17. G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A. Di Filippo, A. Di Matteo, M. Colaneri, Modelling the covid-19 epidemic and implementation of population-wide interventions in Italy. Nat. Med. **26**(6), 855–860 (2020)
18. D.F. Gudbjartsson, A. Helgason, H. Jonsson, O.T. Magnusson, P. Melsted, G.L. Norddahl, J. Saemundsdottir, A. Sigurdsson, P. Sulem, A.B. Agustsdottir, et al., Spread of sars-cov-2 in the Icelandic population. New Engl. J. Med. **382**, 2302–2315 (2020)
19. X. He, E.H. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y.C. Lau, J.Y. Wong, Y. Guan, X. Tan, et al., Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat. Med. **26**(5), 672–675 (2020)
20. M.J. Keeling, K.T. Eames, Networks and epidemic models. J. Royal Soc. Interface **2**(4), 295–307 (2005)
21. W.O. Kermack, A.G. McKendrick, A contribution to the mathematical theory of epidemics. Proc. Royal Soc. Lond. A **115**(772), 700–721 (1927)
22. D. Kim, A. Quaini, Coupling kinetic theory approaches for pedestrian dynamics and disease contagion in a confined environment. Math. Models Methods Appl. Sci. **30**, 1893–1915 (2020)
23. C. Kipnis, C. Landim, *Scaling Limits of Interacting Particle Systems* (Springer, Berlin, 1999)
24. D.A. Knopoff, F.A. Trucco, A compartmental model for antibiotic resistant bacterial infections over networks. Int. J. Biomath. **13**(1), 2050001 (2020)
25. E. Lavezzo, E. Franchin, C. Ciavarella, G. Cuomo-Dannenburg, L. Barzon, C. Del Vecchio, L. Rossi, R. Manganelli, A. Loregian, N. Navarin, et al., Suppression of a sars-cov-2 outbreak in the Italian municipality of vo'. Nature **584**(7821), 425–429 (2020)
26. J. Murray, *Mathematical Biology II: Spatial Models and Biomedical Applications*, vol. 3 (Springer, Berlin, 2001)
27. A.S. Perelson, Modelling viral and immune system dynamics. Nat. Rev. Immunol. **2**(1), 28–36 (2002)
28. T.C. Porco, K.A. Holbrook, S.E. Fernyak, D.L. Portnoy, R. Reiter, T.J. Aragón, Logistics of community smallpox control through contact tracing and ring vaccination: a stochastic network model. BMC Publ. Health **4**(1), 34 (2004)
29. H.C. Tuckwell, R.J. Williams, Some properties of a simple stochastic epidemic model of SIR type. Math. Biosci. **208**(1), 76–97 (2007)
30. A. Vespignani, Modelling dynamical processes in complex socio-technical systems. Nat. Phys. **8**(1), 32–39 (2012)
31. A. Zeb, E. Alzahrani, V.S. Erturk, G. Zaman, Mathematical model for coronavirus disease 2019 (COVID-19) containing isolation class. BioMed Res. Int. **2020**, 3452402 (2020)

Productivity in Times of Covid-19: An Agent-Based Model Approach

José F. Fontanari

1 Introduction

The trade-off between productivity at work and social interaction and, in particular, the influence of mood on human cognitive function are long-established research topics of social psychology (see, e.g., [1–3]). But the recent lockdowns and social restriction measures aiming at slowing down the spread of the SARS-CoV-2 have placed these themes at the forefront of public attention, with mental health pundits predicting a reduction in overall productivity during this time, despite the quick switch to remote work.

In fact, a popular explanation for the productivity reduction in a remote work scenario is the so-called water-cooler effect which is based on the anecdotal observations that placing an inanimate object for people to congregate around can stir up casual conversations and, consequently, increase company productivity [4]. Of course, this is not obvious at all, since for a singular worker the extra downtime communicating may actually detract from its productivity as a whole. Fortunately, the economic literature offers more solid explanations for the positive correlation between social interaction or communication among workmates and productivity [5]. For instance, the peer pressure that generates feelings of guilt or shame when one's performance falls short of the expectation may increase focus on the work and hence increase productivity [6, 7]. In addition, it is through social interactions that workmates learn from each other and build up skills that they otherwise would not have. Interestingly, this so-called knowledge spillover effect is the usual explanation for the strong link between economic growth and the concentration of people in cities [8, 9].

J. F. Fontanari (✉)
Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, SP, Brazil
e-mail: fontanari@ifsc.usp.br

N. Bellomo, M. A. J. Chaplain (eds.), *Predicting Pandemics in a Globally Connected World, Volume 1*, Modeling and Simulation in Science, Engineering and Technology, https://doi.org/10.1007/978-3-030-96562-4_7

However, since aside vaccination the main measure to curb the spread of SARS-CoV-2 is physical distancing, rather than social distancing, one may argue that internet-based and social media usage may mitigate the lack of physical proximity (i.e., face-to-face social interaction within the same physical space) with workmates [10, 11]. It is unclear, however, if use of technology to socialize remotely can significantly improve mood: evidence of heightened psychological problems among the youth in the wake of this pandemic indicates that the abundance of virtual social contacts may have actually little or even negative impact on the feelings of loneliness [12, 13]. The key issue here is, of course, the quality of the social interactions, as the so-called Zoom fatigue illustrates so nicely [14], regardless of whether they are virtual or physical [15]. We refer the reader to Isaac Asimov's novel *The Naked Sun* for an insightful analysis of the (surprising) collateral effects of prolonged physical distancing [16].

Studying the interplay between mood, social behavior, and productivity using a quantitative approach is particularly important at this time of crisis when physical distancing and quarantining are enforced all over the world. In this perspective, we envisage a scenario where the mathematical and computational studies of the Covid-19 pandemics will account not only for the transmission of the physical infection in the distinct age and socioeconomic segments of the population (see [17] for an in-depth study of this issue) but also for the effects of the measures to counter that transmission on the mental health of the population. Here, it is opportune to draw attention to the fact that for many people the choice is between physical infection and mental breakdown [18].

Accordingly, we have recently initiated a research program aiming at studying the effects of physical distancing on mood and productivity using agent-based models of a workplace scenario [19, 20]. In this approach, the mood (e.g., motivation to work) of an agent is measured by a real variable that determines the propensity of the agent to initiate a social interaction. So, on the one hand, we assume that mood determines behavior. On the other hand, we assume that mood improves if the agent is socializing and deteriorates if the agent is working, so that behavior determines mood. The instantaneous productivity is a function of both behavior and mood and, in particular, we assume that no work is done during socialization in harmony with the "water-cooler" scenario and that productivity increases as mood improves [3]. An agent can either socialize (i.e., talk to a workmate) or work. These are the only types of behavior considered in our model. The feedback between mood and behavior is responsible for the nontrivial results of the work group dynamics that we review and develop in this chapter.

The main outcome of our model is the finding that the decrease of the frequency, quality, or duration of social contacts leads the agents to enter a regime of burnout in which their motivation to work tends to negative infinity as the duration of the workday increases. In this regime, the time-averaged cumulative productivity tends to zero, as expected. The regime of mental breakdown is separated from a healthy regime, where the motivation to work is finite, by a continuous phase transition in the sense that the fraction of working agents in the group changes continuously when transitioning between those regimes. In the healthy regime, we

find that the time-averaged cumulative productivity increases with the frequency and quality of the social contacts, in accordance with the water-cooler scenario presumption. The unexpected threshold phenomenon leading to a low-productivity burnout situation highlights our unfamiliarity with the mental health consequences of physical distancing.

The rest of this chapter is organized as follows. In Sect. 2, we present the rules that govern the behavior of the agents given their motivations to work, as well as the manner that those motivations are affected by their behaviors. In Sect. 3, we offer an analytical (mean-field) approximation that describes the agent-based model using three difference equations that couple the mean fraction of working agents in the group, the mean motivation to work of a typical agent, and the mean cumulative productivity per agent. These equations allow the identification and characterization of the healthy and burnout working regimes mentioned above. In Sect. 4, we discuss the setting of the several functions and parameters introduced in the definition of the model and whose specifications are necessary to run the simulations. In Sect. 5 we present the results of the simulations of the agent-based model together with the predictions of the mean-field approximation. We find a good quantitative agreement between these approaches. Finally, in Sect. 6 we summarize our main results and present some concluding remarks.

2 Model

In this section we offer a model of the dynamics of a work group composed of N agents that interact socially and work together during a period of duration T. For instance, the typical duration of a workday is $T = 480$ minutes. We assume that an agent is more productive when it is motivated (i.e., in a good mood to work) [3, 21, 22], and its motivation improves when it engages in some form of social interaction [1, 23]. In this perspective, motivation has an emotional dimension, since talking to its peers improves the mood of the agent and helps it to focus on the task, as well as an informational dimension, since talking to peers may result in the acquisition of information that can aid the agent to consummate its task.

The motivation of an agent, say agent k, at time t is measured by the parameter $L_k^t \in \mathbb{R}$ with $t = 0, \delta t, \ldots, T$ that, in turn, determines the propensity of this agent to seek and engage in social interaction, as well as to end an ongoing interaction. The time unit δt will be specified later. We assume that low motivated people seek company and high motivated people focus on work. In addition, we assume that L_k^t is affected differently depending on whether agent k is working or talking with its peers. Moreover, we assume that the instantaneous productivity of agent k at time t, denoted by π_k^t, is a function of its motivation $\pi_k^t = \pi(L_k^t)$ if it is working and $\pi_k^t = 0$ if it is talking. This is in line with the experimental findings that being with other people would be associated with poorer cognitive task performance than when being alone [24]. Here we choose $\pi(x)$ as an arbitrary increasing function of its argument so as to ensure that the productivity increases with the motivation of the

agent. We note that in our framework only lone agents are capable of doing work, and their productivities depend on their motivations, which, in turn, can increase only through social interactions. These assumptions produce a complex interplay between motivation, behavior, and productivity that is responsible for the nontrivial results of the group dynamics.

More pointedly, if agent k is working (and so is alone), then the probability that it will attempt to instigate a conversation is given by $p_k = p(L_k^t)$, where $p(x) \in [0, 1]$ is an arbitrary function. When the lone agent k decides to instigate a conversation, it selects a number m of contact attempts, where $m = 0, 1, \ldots$ is a random variable drawn from a Poisson distribution of parameter q. In each contact attempt, a workmate is selected at random among the $N - 1$ agents in the group and, in case the selected agent is alone at that moment, a conversation is initiated. In this case, the agent k halts its search for company. If none of the m selected agents are alone, then the attempt of agent k to socialize fails.

We assume that a conversation involves two agents only and that the lone agent that is approached by agent k is obliged to accept the interaction, regardless of its motivation. This pro-social behavior is chosen in order to not further complicate the model, but it can be justified in terms of workplace social norms [25], especially during the current pandemic when there is a pressure to talk to everyone because one worries that they are lonely and one does not want to turn them down. Of course, this pro-social behavior is one of the causes of the so-called Zoom fatigue.

If agent k is socializing, then the probability that it will unilaterally interrupt the conversation is given by $r_k = r(L_k^t)$, where $r(x) \in [0, 1]$ is another arbitrary function. Of course, in the case the conversation is ended both agent k and its mate resume their work.

Finally, we assume that the rate of change of the motivation of agent k is determined by the function $M_w(L_k^t) \in \mathbb{R}$ if it is alone (i.e., working) and by the function $M_s(L_k^t) \in \mathbb{R}$ if it is socializing (i.e., not working).

The asynchronous dynamics of the work group at time t proceeds as follows. In the time interval δt, we pick an agent at random, say agent k, and check if it is working or socializing. In case it is working, we change its motivation L_k and cumulative productivity Π_k following the prescriptions

$$L_k^{t+\delta t} = L_k^t + M_w(L_k^t)\delta t \tag{1}$$

and

$$\Pi_k^{t+\delta t} = \Pi_k^t + \pi_k^t \delta t. \tag{2}$$

Then we test if agent k will attempt to initiate a conversation using the socializing probability $p_k = p(L_k^t)$. As mentioned before, this attempt involves the selection with replacement of at most m workmates until another lone agent is found. In case agent k is socializing, we change its motivation according to the prescription

$$L_k^{t+\delta t} = L_k^t + M_s(L_k^t)\delta t, \tag{3}$$

whereas its cumulative productivity remains unaltered, i.e., $\Pi_k^{t+\delta t} = \Pi_k^t$. Then we check if it will terminate the conversation using the termination probability $r_k = r(L_k^t)$. In case it does, both agent k and its mate resume work at time $t+\delta t$. As usual in such asynchronous update scheme, we choose the time increment as $\delta t = 1/N$ so that during the increment from t to $t+1$ exactly N, though not necessarily distinct, agents are chosen to follow the update rules.

In order to clarify possible ambiguities of the above informal account of the asynchronous group dynamics, it is convenient to describe it in a more formal manner. In the following, we will denote the state working (or alone) by the symbol w and the state socializing by the symbol s.

Given that agent k is working at time t, the probability that it will continue so at time $t+\delta t$ is

$$\begin{aligned} Q_k(w, t+\delta t \mid w, t) = {} & \frac{1}{N}\left[(1-p_k) + p_k \exp\left(-q\frac{N_w^t - 1}{N-1}\right)\right] \\ & + \frac{1}{N}\sum_{i\in\mathcal{L}_w^t; i\neq k}\left[(1-p_i) + p_i \exp\left(-\frac{q}{N-1}\right)\right] \\ & + \frac{N - N_w^t}{N}, \end{aligned} \tag{4}$$

where N_w^t and $N - N_w^t$ are the numbers of working and socializing agents at time t, respectively. The sum in the second term of the rhs of this equation is over the subgroup of working agents $\mathcal{L}_w^t$, except agent k, at time t. For notational simplicity, we have omitted the time dependence of p_k. The first term of the rhs of Eq. (4) accounts for the possibility that agent k is the agent selected for update, which is an event that happens with probability $1/N$. In this case there are two possibilities: agent k decides to continue working, which happens with probability $1 - p_k$ or decides to instigate a conversation but fails to find another lone agent, which happens with probability

$$p_k \sum_{m=0}^{\infty} e^{-q}\frac{q^m}{m!}\left(1 - \frac{N_w^t - 1}{N-1}\right)^m = p_k \exp\left(-q\frac{N_w^t - 1}{N-1}\right). \tag{5}$$

The second term of the rhs of Eq. (4) accounts for the possibility that a working agent $i \neq k$ is chosen for update and that this agent either decides to continue working, which has probability $1 - p_i$, or instigate a conversation with any other agent but agent k, which has probability

$$p_i \sum_{m=0}^{\infty} e^{-q}\frac{q^m}{m!}\left(1 - \frac{1}{N-1}\right)^m = p_i \exp\left(-\frac{q}{N-1}\right). \tag{6}$$

Finally, the third term of the rhs of Eq. (4) accounts for the possibility that the agent selected for update in the time interval δt is one of the $N - N_w^t$ agents that are socializing at time t. Since a working agent at time t can either continue working or start socializing at time $t+\delta t$, the probability that the working agent k at time t starts socializing during the time interval δt is readily obtained from the complement rule of probability:

$$Q_k(s, t+\delta t \mid w, t) = \frac{p_k}{N}\left[1 - \exp\left(-q\,\frac{N_w^t - 1}{N-1}\right)\right] + \frac{1}{N}\sum_{i\in\mathcal{L}_a^t; i\neq k} p_i\left[1 - \exp\left(-\frac{q}{N-1}\right)\right]. \tag{7}$$

Next, we assume that agent k is socializing with agent k' at time t. The probability that this interaction continues during the time interval δt is simply

$$Q_{k,k'}(s, t+\delta t \mid s, t) = \frac{1}{N}(1 - r_k) + \frac{1}{N}(1 - r_{k'}) + \frac{N-2}{N}, \tag{8}$$

where we have omitted the time dependence of r_k and $r_{k'}$. Here the first two terms of the rhs of this equation account for the events that agents k and k' are selected for update and they choose not to interrupt their conversation. The last term of the rhs of Eq. (8) accounts for the event that any other agent, aside from k and k', is selected for update at time t. As before, the event that k and k' will terminate their conversation during the time increment δt is complementary to the event that they will continue their conversation, i.e.,

$$Q_{k,k'}(w, t+\delta t \mid s, t) = \frac{1}{N}\left(r_k + r_{k'}\right). \tag{9}$$

At this stage, it is instructive to note that Eqs. (4) and (8) are probabilities of events that occur in the time interval δt and so they should be proportional to δt, which is in fact the case since we set $\delta t = 1/N$, as pointed out before. We stress, however, that we will not consider the unrealistic limit of infinitely large work groups $N \to \infty$, which would correspond to a continuous-time model of the group dynamics. In addition, we note that Eq. (8) introduces a short-time correlation between agents k and k' that hinders an exact analytical approach to solve the model. However, in the next section we offer a simple mean-field approximation that yields a fairly good description of the macroscopic features of the group dynamics.

3 Mean Field Approximation

Clearly, a macroscopic quantity of interest is the number of lone agents N_w^t in the group at time t. This quantity is a random variable that can increase by two units, decrease by two units, or remain the same during the time interval δt. More pointedly, given N_w^t and the motivations L_k^t for $k = 1, \ldots, N$, the probabilities of those three events are

$$P\left(N_w^{t+\delta t} = N_w^t + 2\right) = \frac{1}{N} \sum_{k \in \mathcal{L}_s^t} r_k, \tag{10}$$

$$P\left(N_w^{t+\delta t} = N_w^t - 2\right) = \frac{1}{N} \sum_{k \in \mathcal{L}_w^t} p_k \left[1 - \exp\left(-q \frac{N_w^t - 1}{N - 1}\right)\right], \tag{11}$$

and $P\left(N_w^{t+\delta t} = N_w^t\right) = 1 - P\left(N_w^{t+\delta t} = N_w^t + 2\right) - P\left(N_w^{t+\delta t} = N_w^t - 2\right)$. Here $\mathcal{L}_s^t$ stands for the subgroup of socializing agents at time t. Hence the expected number of working agents at time $t + \delta t$ given that there are N_w^t working agents at time t is

$$\langle N_w^{t+\delta t} \rangle = N_w^t + 2\left[P\left(N_w^{t+\delta t} = N_w^t + 2\right) - P\left(N_w^{t+\delta t} = N_w^t - 2\right)\right]. \tag{12}$$

Following the same reasoning, we can write the expected motivation of agent k at $t + \delta t$ as

$$\begin{aligned} \langle L_k^{t+\delta t} \rangle &= \left[L_k^t + M_w(L_k^t)\delta t\right] \frac{1}{N} \frac{N_w^t}{N} + \left[L_k^t + M_s(L_k^t)\delta t\right] \frac{1}{N} \frac{N - N_w^t}{N} + L_k^t \frac{N-1}{N} \\ &= L_k^t + \left[\frac{N_w^t}{N}\left(M_w(L_k^t) - M_s(L_k^t)\right) + M_s(L_k^t)\right] \frac{\delta t}{N}, \end{aligned} \tag{13}$$

where we have used that the probabilities that agent k is working or socializing at time t are N_w^t/N and $(N - N_w^t)/N$, respectively.

Finally, given the cumulative productivity of agent k at time t, we can write the expected cumulative productivity at $t + \delta t$ as

$$\begin{aligned} \langle \Pi_k^{t+\delta t} \rangle &= \left[\Pi_k^t + \pi(L_k^t)\delta t\right] \frac{1}{N} \frac{N_w^t}{N} + \Pi_k^t \frac{1}{N} \frac{N - N_w^t}{N} + \Pi_k^t \frac{N-1}{N} \\ &= \Pi_k^t + \left[\frac{N_w^t}{N} \pi(L_k^t)\right] \frac{\delta t}{N}. \end{aligned} \tag{14}$$

To proceed further we make the usual mean-field assumption $N_w^t \approx \langle N_w^t \rangle$, $L_k^t \approx \langle L_k^t \rangle$, and $\Pi_k^t \approx \langle \Pi_k^t \rangle$ (see, e.g., [26]). In addition, we assume that the mean motivation and the mean cumulative productivity are the same for all agents, i.e., $\langle L_k^t \rangle = \langle L^t \rangle \equiv l^t$ and $\langle \Pi_k^t \rangle = \langle \Pi^t \rangle \equiv \rho^t$. Moreover, we introduce the notation $\eta^t = \langle N_w^t \rangle / N$ for the fraction of working agents at time t. These assumptions suffice for writing the mean-field version of the group dynamics, viz.,

$$\eta^{t+\delta t} = \eta^t + 2\left[(1-\eta^t)r(l^t) - \eta^t p(l^t)\left(1 - \exp\left(-q\frac{\eta^t - 1/N}{1-1/N}\right)\right)\right]\delta t, \quad (15)$$

$$l^{t+\delta t} = l^t + \left[\eta^t\left(M_w(l^t) - M_s(l^t)\right) + M_s(l^t)\right]\frac{\delta t}{N}, \quad (16)$$

and

$$\rho^{t+\delta t} = \rho^t + \eta^t \pi(l^t)\frac{\delta t}{N}, \quad (17)$$

where we have used $\delta t = 1/N$ in Eq. (15) to stress the incremental nature of the intensive variable η^t. We note that the mean cumulative productivity has no influence on the mean motivation l^t or on the fraction of working agents η^t. Of course, a possible direction to improve the model is to allow the productivity to influence the motivation and the propensity of the agents to instigate a conversation in the workplace.

In order to better appreciate the nuances of our model and to constrain the several arbitrary functions introduced in its setup, we consider now the time asymptotic limit $T \to \infty$ of the mean-field equations. We stress, however, that since T is the duration of a workday, it must be finite in a realistic scenario. In the case that the group dynamics converges to a fixed point $l^{t+\delta t} = l^t = l^*$ in the asymptotic limit, the equilibrium fraction of working agents $\eta^{t+\delta t} = \eta^t = \eta_h^*$ is given by:

$$\eta_h^* = \frac{M_s(l^*)}{M_s(l^*) - M_w(l^*)} \quad (18)$$

with l^* given by the solution of the transcendental equation

$$-\frac{M_w(l^*)r(l^*)}{M_s(l^*)p(l^*)} = 1 - \exp\left(-q\frac{\eta_h^* - 1/N}{1-1/N}\right). \quad (19)$$

The subscript h in our notation for the equilibrium fraction of working agents η_h^* stands for healthy since l^* is finite in this equilibrium and so it characterizes a healthy work regime. The condition $\eta_h^* \in [0, 1]$ requires that either $M_w(l^*) < 0$ and $M_s(l^*) > 0$ or $M_w(l^*) > 0$ and $M_s(l^*) < 0$. Since l^t measures the motivation of a generic agent, here we will assume that $M_w(l^t) < 0$ and $M_s(l^t) > 0$, which, according to Eqs. (1) and (3), means that the motivation of an agent increases when it is socializing and decreases when it is working (i.e., when it is alone).

An interesting situation occurs when Eq. (19) has no solution so that $\lim_{t\to\infty} l^t \to -\infty$. This divergence characterizes a burnout regime where the equilibrium fraction of working agents η_b^* is given by the solution of the equation

$$\lim_{l^t \to -\infty} \frac{r(l^t)}{p(l^t)} = \frac{\eta_b^*}{1-\eta_b^*} \left[1 - \exp\left(-q \frac{\eta_b^* - 1/N}{1 - 1/N} \right) \right], \tag{20}$$

which is obtained from Eq. (15) by setting $\eta^{t+\delta t} = \eta^t = \eta_b^*$ and the subscript b in η_b^* stands for burnout.

Regardless of the working regime, the cumulative mean productivity (17) always diverges in the time asymptotic limit, as expected, since the instantaneous productivity of a working agent is non-negative. Hence, in order to get useful information about the productivity of the agents we introduce the time-averaged cumulative productivity per agent

$$\bar{\rho}_T = \rho^T / T. \tag{21}$$

Unfortunately, we cannot derive an analytical expression for this quantity in the limit $T \to \infty$, so we need to resort to the numerical iteration of the system of Eqs. (15)–(17) to obtain information about the productivity of the work team. However, before embarking on the numerical analysis of those difference equations and on the simulations of the agent-based model we need to finish the setup of our model.

4 Setting the Model Functions

Up to now we have left unspecified the functions $p(l)$, $r(l)$, and $\pi(l)$ that determine the effect of the motivation l on the behavior and productivity of the agents, as well as the functions $M_w(l) < 0$ and $M_s(l) > 0$ that, in turn, determine how the behavior of the agents influences their motivations. In order to simulate the model, we need to specify those functions and in this section we propose particular functional forms for them.

We begin by assuming that the probability to instigate a conversation is a decreasing function of the motivation of the agents:

$$p(l) = \frac{1}{2} \left[1 - \tanh(\beta l) \right], \tag{22}$$

where $\beta \geq 0$ is a parameter that determines the influence of the motivation on the behavior of the agent. For instance, for $\beta = 0$, the motivation has no effect on an agent's decision to instigate or not a conversation, whereas for $\beta \to \infty$ a working agent will always attempt to socialize and hence stop working when $l < 0$. The point of the choice (22) is that it guarantees that high-motivated agents will be unlikely to seek to socialize and that low-motivated agents will actively seek to instigate conversations. Moreover, we assume that the probability that a socializing agent terminates a conversation does not depend on its motivation, i.e., $r(l) = r \in [0, 1]$, since there are many external factors that may result in the interruption of

a conversation, in contrast to the longing to socialize, which is most likely fed by internal factors [27].

We recall that the instantaneous productivity of an agent is nonzero only in the case the agent is working. In this case, we assume that $\pi(l)$ is an increasing, non-negative function of the motivation l with $\lim_{l\to-\infty}\pi(l) = 0$ and $\lim_{l\to\infty}\pi(l) = c$. The limiting value c is immaterial since it gives the scale of the cumulative productivity (see Eqs. (2) and (17)) only and so we can set $c = 1$ without loss of generality. A simple choice that satisfies those constraints is

$$\pi(l) = \frac{1}{2}\left[1 + \tanh(\alpha l)\right], \tag{23}$$

where $\alpha \geq 0$ is a parameter that determines the influence of the motivation on the productivity of working agents.

Finally, to complete the setup of our model of group dynamics, we assume, for the sake of simplicity, that the rates of change of the motivation are constant, i.e., $M_w(l) = -w < 0$ and $M_s(l) = s > 0$. This means that the motivation decreases by a factor $w\delta t$ when the agent is working and that the motivation increases by a factor $s\delta t$ when the agent is socializing. We note that we can set $w = 1$ without loss of generality, since this parameter can be removed from our equations by a proper rescaling of L_k, s, β and α.

With the above choices, we can rewrite Eqs. (18) and (19) and obtain explicit expressions for η_h^* and l^*, viz.,

$$\eta_h^* = \frac{s}{1+s}, \tag{24}$$

$$l^* = -\frac{1}{2\beta}\ln\left(\frac{\Lambda}{1-\Lambda}\right), \tag{25}$$

where

$$\Lambda = \frac{r/s}{1 - \exp\left(-q\,\frac{s/(1+s)-1/N}{1-1/N}\right)}. \tag{26}$$

This fixed point exists provided that $\Lambda < 1$ and a necessary (but not sufficient) condition for this happening is $r/s < 1$. In fact, a small value of r implies that the conversations last longer and a large value of s implies that they bring about a substantial increase of motivation. (We recall that the comparison baseline of s is the decrement of motivation of working agents, viz., $w = 1$.) Hence, the lesser the rate r/s, the more motivated the agents are, provided, of course, that they can find conversation partners whenever they need one. The parameter q is key to the success or failure of the search for those partners.

In the case that $\Lambda \geq 1$ we find that $l^t \to -\infty$ in the limit $t \to \infty$, i.e., iteration of Eqs. (15) and (16) with $\delta t = 1/N$ (see, e.g., Fig. 1) does not lead to a fixed point

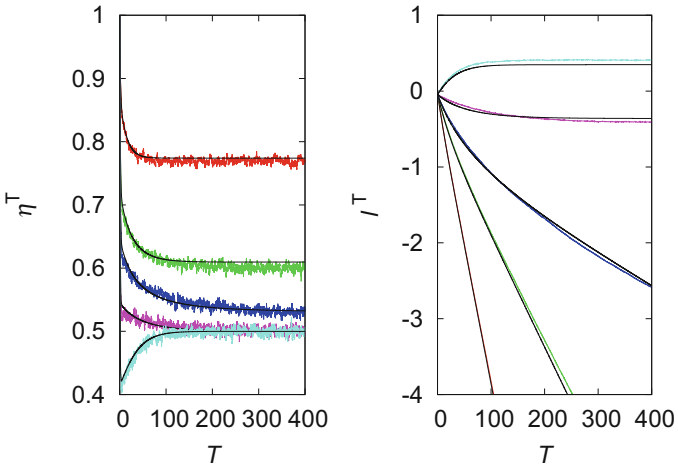


Fig. 1 Mean fraction of working agents η^T (left panel) and mean motivation per agent l^T (right panel) at the end of a workday of duration T for a group of size $N = 15$. The mean number of contact attempts is (from top to bottom in the left panel and from bottom to top in the right panel) $q = 0.1, 0.3, 0.5, 1$, and 3. The other parameters are $r = 0.25$, $s = 1$, $\beta = 1$, and $\alpha = 1$. The mean-field prediction for the critical point is $q_c = 0.587$. The colored lines are the averages over 10^3 independent agent-based simulations and the black lines are the mean-field predictions

for the variable l^t. However, it does for the variable η^t that tends to the finite value η_b^* given by Eq. (21), which reduces to

$$r = \frac{\eta_b^*}{1-\eta_b^*}\left[1 - \exp\left(-q\,\frac{\eta_b^* - 1/N}{1-1/N}\right)\right], \tag{27}$$

since $\lim_{t\to\infty} p(l^t) = 1$ and $r(l^t) = r$. We note that for $\Lambda = 1$ Eq. (27) reduces to Eq. (24), i.e., $\eta_h^* = \eta_b^*$, so that the transition between the healthy and burnout working regimes is continuous regarding the asymptotic mean fraction of working agents.

The condition $\Lambda = 1$ determines the critical values of the control parameters of the model. It is convenient at this stage to select the leading parameter of our analysis, whose variation best reflects the effects of the physical distancing measures on the work routine. The natural choice here is the mean number of attempts to make a social contact q. Regardless of whether the contact attempts are physical or remote, their overall number diminished significantly during the pandemics as, for instance, the causal encounters at the water cooler or at the cafeteria were no longer possible. Hence the threshold value that separates the healthy and the burnout working regimes is

$$q_c = -\frac{1-1/N}{s/(1+s) - 1/N}\ln(1 - r/s) \tag{28}$$

with $r/s < 1$. The healthy regime occurs for $q > q_c$ (i.e., $\Lambda < 1$) and the burnout regime for $q \leq q_c$ (i.e., $\Lambda \geq 1$). In the case that $r/s > 1$, the model exhibits the burnout regime only. In this case, the equilibrium fraction of lone agents η_b^* given by Eq. (27) does not depend on s.

5 Simulations

Here we present the results of the agent-based simulations as well as the theoretical predictions of the mean-field approximation. Throughout this section, we set $N = 15$, which is not only the typical number of close friends, i.e., people one trusts and spends time with regularly [28], but also a plausible size for a work team. In addition, we set the initial conditions of each run as $N_w^0 = N$ and $L_k^0 = \Pi_k^0 = 0$ for $k = 1, \ldots, N$. In terms of the mean-field variables, these conditions amount to the choices $\eta^0 = 1$, $l^0 = 0$, and $\rho^0 = 0$.

In Fig. 1, we show the fraction of working agents and the mean motivation per agent at the end of a workday of duration T. It is clear from these results that the agent-based model also exhibits a critical value of q below which the mean motivation diverges to negative infinity (burnout regime) and above which converges to a positive or negative finite value (healthy regime). We recall that in the previous sections we have discussed this transition using the mean-field results only. In fact, the mean-field predictions describe the simulation results quite well with a few perceptible discrepancies that the scale of the figure has obscured somewhat. We note that, in accordance with the mean-field prediction Eq. (24), η^T is not affected by the value of q in the healthy working regime.

Since the phase transition between the healthy and burnout regimes that takes place in the limit $T \to \infty$ is the main outcome of our workplace model, in Fig. 2 we show the simulation results for a large value of T together with the mean field predictions for the equilibrium. The point here is to highlight the characteristics of the distinct working regimes. Despite some minor quantitative discrepancies, the mean-field approximation describes the simulation results and, in particular, the location of the transition point rather well. We draw attention to the curious finding that the rate of increment of motivation due to socialization s has no influence on the number of working agents in the burnout regime. The limit $q \to \infty$ guarantees that a working agent will always find a conversation partner if there is one available. In this case, the mean-field approximation yields $\eta_h^* = s/(1+s)$ and $l^* = -(1/2\beta)\ln[(r/s)/(1-r/s)]$ if $r/s < 1$, and $\eta_b^* = r/(1+r)$ and $l^* \to -\infty$ if $r/s \geq 1$.

It is instructive to look into the predictions of the mean-field approximation near the critical point q_c for $r/s < 1$. In the healthy working regime ($q > q_c$) we find

$$l^* \approx -\frac{1}{2\beta}\ln(q - q_c) \tag{29}$$

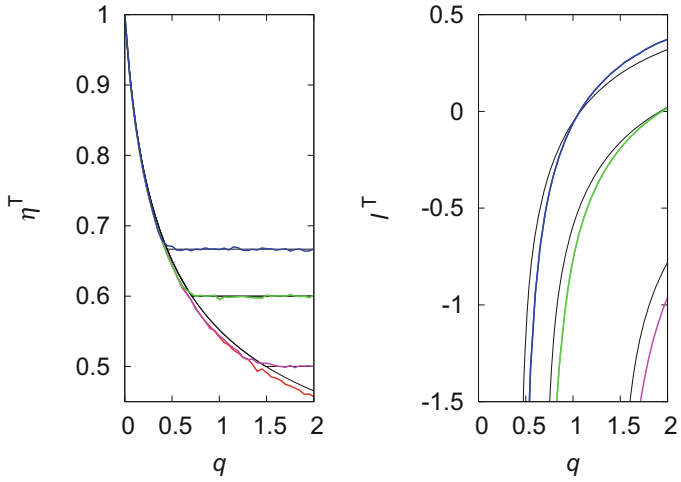


Fig. 2 Mean fraction of working agents η^T (left panel) and mean motivation per agent I^T (right panel) at the end of a workday of duration $T = 2000$ as functions of the mean number of contact attempts q for (from top to bottom in both panels) $s = 2$, 1.5, 1, and 0.5. The other parameters are $N = 15$, $r = 0.5$, $\beta = 1$, and $\alpha = 1$. The colored lines are the averages over 10^4 independent agent-based simulations and the black lines are the equilibrium mean-field predictions. The data for $s = 0.5$ is not shown in the right panel because I^* diverges to negative infinity in the mean-field approximation and the simulations yield results outside the range of the y-axis

and $\eta_h^* = s/(1+s)$, whereas in the burnout regime ($q < q_c$) we find

$$\eta_b^* \approx \frac{s}{1+s} + \mathcal{A}(1 - \frac{q}{q_c}), \tag{30}$$

where

$$\mathcal{A} = -\ln(1 - r/s)\frac{s(s-r)(1-1/N)}{q_c s(s-r) + r(1+s)^2(1-1/N)} > 0. \tag{31}$$

Hence, if we define the order parameter of the phase transition as $\delta = \eta_b^* - \eta_h^*$, then $\delta \sim (q_c - q)$ as we approach the critical point from the burnout regime. Thus, the phase transition separating the healthy and the burnout working regimes is a continuous phase transition regarding the order parameter δ.

We turn now to the study of the productivity of the agents and in Fig. 3 we show the time-averaged cumulative productivity per agent $\bar{\rho}_T$ as function of the duration of the workday T. In this case, the discrepancies between the mean-field predictions and the simulations are very noticeable. Although in the case of large values of q the agreement is very good, the mean-field approximation underestimates $\bar{\rho}_T$ in the other cases. The analytical approximation does a good job at the qualitative level, however. As T increases, we find that $\bar{\rho}_T$ tends to zero with $1/T$ in the burnout regime, since for large negative values of the motivation variable I^t the

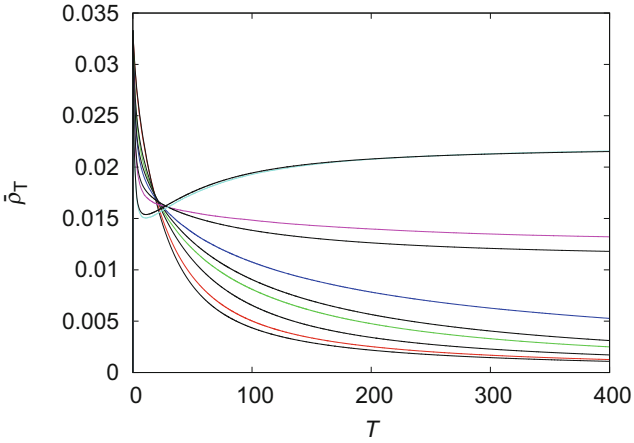


Fig. 3 Time-averaged cumulative productivity per agent $\bar{\rho}_T$ at the end of a workday of duration T for a group of size $N = 15$. The mean number of contact attempts is (from bottom to top at $T = 400$) $q = 0.1, 0.3, 0.5, 1$, and 3. The other parameters are $r = 0.25$, $s = 1$, $\beta = 1$, and $\alpha = 1$. The colored lines are the averages over 10^3 independent agent-based simulations and the black lines are the mean-field predictions. Those lines are indistinguishable for $q = 3$, otherwise the mean-field results always underestimate the simulation results

instantaneous productivity $\pi(l^t)$ (see Eq. (23)) is vanishingly small. However, $\bar{\rho}_T$ tends to constant value in the healthy working regime. Of course, the relevant case is that for which the duration of the workday T is finite. In this case, we note that the cumulative productivity increases with increasing q provided T is not too small. This is not a straightforward conclusion since increase of q increases socialization and hence decreases the number of agents that are actually working. Interestingly, for very short workdays, say $T < 30$, this conclusion is reversed: increase of q results in a slight decrease of productivity. This is actually expected since for such short workdays the decrease of productivity during a conversation will not be compensated later with the boost of motivation resulting from the conversation. Finally, since we have set $l^0 = \rho^0 = 0$ and $\eta^0 = 1$, we have $\bar{\rho}_{\delta t} = \pi(0)/N = 1/(2N)$, and so $\bar{\rho}_T$ exhibits a discontinuity at $T = 0$. In addition, the maximum time-averaged cumulative productivity per agent is $\bar{\rho}_T = 1/N$, which corresponds to the unattainable situation where the agent works with maximum motivation during the entire workday. In this case the time-averaged cumulative productivity of the group is $N\bar{\rho}_T = 1$.

We note that the time-averaged cumulative productivity per agent decreases with increasing T in the burnout regime because the motivation to work becomes more and more negative as the duration of the workday T increases, as illustrated in the right panel of Fig. 1. We recall that we set the initial motivation at the neutral value $L_k = 0$ for all k. In addition, $\bar{\rho}_T$ may also decrease with increasing T in the healthy regime if l^T tends to a negative value in the limit $T \to \infty$, which is the case of

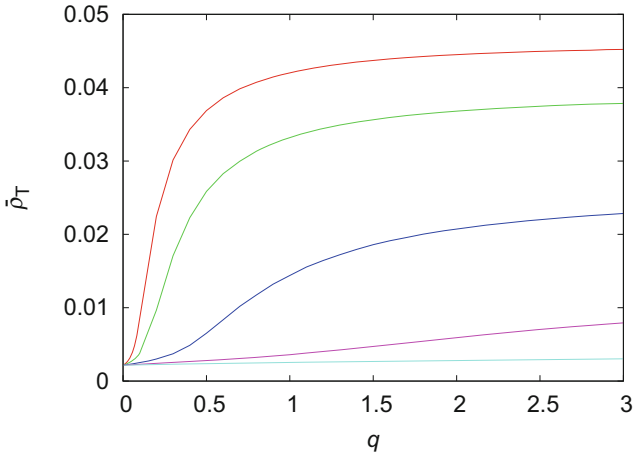


Fig. 4 Time-averaged cumulative productivity per agent $\bar{\rho}_T$ at the end of a workday of duration $T = 400$ for a group of size $N = 15$ as function of the mean number of contact attempts q for (from top to bottom) $s = 3, 2, 1, 0.5$, and 0.25. We recall that s is a measure of the quality of the social interaction. The other parameters are $r = 0.25$, $\beta = 1$, and $\alpha = 1$. The lines are the averages over 10^3 independent agent-based simulations

$q = 1$. However, if that value is positive, as in the case of $q = 3$, then we find that $\bar{\rho}_T$ actually increases with T for large T.

To better appreciate the effects on the productivity of the two parameters of the model that are more clearly influenced by the physical distancing measures, viz. q and s, in Fig. 4 we show $\bar{\rho}_T$ evaluated at $T = 400$. The results are similar to those obtained at equilibrium (i.e., for $T \to \infty$), except that $\bar{\rho}_T \sim 1/T = 0.0025$ in the burnout regime, but the choice of a finite duration of the workday is more realistic. We have omitted the mean-field predictions to avoid cluttering the figure. We note that in the burnout regime, $\bar{\rho}_T$ is not affected by the parameter q. This regime occurs for $s \leq r$ and for $q < q_c$ with q_c given by Eq. (28) if $s > r$. The results highlight that the properly averaged cumulative productivity per agent increases with both the possibility of socialization, which is measured by q, and the quality of the social interaction, which is measured by s. These findings agree with the widely publicized reports that loneliness has an extensive negative impact on the economy [29]. As already pointed out, the surprising prediction of our model is that regardless of how likely the possibility of socialization q is, if the quality of the socialization s is lower than a threshold value (r, in the case), then the productivity of the agents will be negligibly small. As discussed before, an analogous threshold phenomenon occurs if we fix $s > r$ and decrease q. Hence, similarly to the epidemiological thresholds that determine the spreading of infectious diseases [30], the very measures aimed to contain those diseases may result in the crossing of the thresholds that lead to mental breakdown (burnout syndrome).

6 Conclusion

Before the Covid-19 pandemic, the World Health Organization had already declared social disconnection a major public health challenge, since the socially isolated face heightened morbidity and mortality risks [31]. In fact, prompted by a report that about 13% of its population feel lonely some or all of the time and that this social disconnection may be costing its economy 32 billion pounds a year [29], the UK created a Ministry of Loneliness in 2018 to address this crisis. Against this current, however, the Covid-19 pandemic has brought unprecedented efforts to enforce physical distancing and quarantining measures all over the world. The research reported in this chapter is part of a research program aiming at addressing from a mathematical and computational perspective the socio-economic implications of those measures.

Our main finding is that decrease of the frequency, quality, or duration of social contacts leads the agents to enter a regime of burnout characterized by the divergence to negative infinity of the motivation to work with the consequent vanishing of productivity. The fact that the fraction of working (lone) agents, which is the only observable quantity in our model, varies continuously at the transition point makes it nearly impossible to perceive the proximity to this regime of mental breakdown.

Another relevant finding is that in the healthy working regime, where the motivation to work is finite, the decrease of the frequency or the quality of the social interactions results in a productivity decrease, although no work is done while the agents are socializing. This finding agrees with the popular expectancy of a positive correlation between communication among workmates and productivity, which is explained in the economic literature by the water-cooler effect, peer pressure, and knowledge spillover [4, 5]. There are, however, some rare reports of increased productivity in remote work experiments [32], which can also be studied (and explained) within our framework by introducing stereotypes that differ in their propensities to initiate a social interaction [19].

Throughout this chapter, we use the expression physical distancing rather than social distancing because it can be argued that the use of technology to socialize remotely may be a substitute for the old-fashioned face-to-face interactions [33]. As pointed out before, there are evidences that this is not the case [12], but, nevertheless, it is a good argument against the claim that the frequency of social interactions, measured by the parameter q in our model, has decreased during the Covid-19 pandemic. However, the quality of the social interactions during the pandemic has most certainly decreased as attested by the Zoom fatigue and many other anecdotal similar phenomena. Our model takes this point into account through the parameter $s > 0$ that measures the efficacy of the social interactions in improving the mood of the agents. In fact, even if the number of contact attempts is unlimited (i.e., $q \to \infty$) and the group size is very large (i.e., $N \to \infty$), an agent can experience burnout in the case that $s < r$, where r is the probability that the agent ends the social interaction.

A word is in order about the prospects for future research. First, we note that although we have stated that our agents can exhibit two behaviors only (viz., working and socializing), there is actually a third behavior, viz., seeking a conversation partner, which we have clumped with the working behavior, for the sake of simplicity. In fact, whereas it is reasonable to assume that the motivation to work decreases at the same rate when the agent is working or searching for a partner, the instantaneous productivity should be starkly different in these two situations, a point that our model fails to consider. Of course, this issue has no effect on the onset of the phase transition, but it can make the model somewhat more realistic regarding the productivity measure. Second, another move in the direction of more realistic scenarios is to consider specific, real-world problems to be solved by the work team and relate productivity with the efficiency of the team to crack them as in the study of distributed cooperative problem-solving systems [34–36]. In this context, the motivation variable has only an informational dimension (i.e., knowledge spillover), since the interaction with workmates allows the agents to copy partial solutions through social or imitative learning [37].

In sum, our agent-based model builds on a variety of experimental findings of social psychology on the interplay between mood, social interaction, and productivity to advance a quantitative explanation for the widespread reports of decreased productivity in the present-day pandemics. Since those studies are correlational in nature, they do not specify quantitative relationships between mood and behavior or between mood and productivity that are necessary to implement a simulation of the workplace scenario. Providentially, we have developed a simple mean-field theory that approximates the simulation results very well so that our framework can be easily adapted to describe different choices for those relationships. However, we stress that rather than offering a precise quantitative characterization of the workplace dynamics, our model offers a flexible framework to study the consequences of isolation and loneliness resulting from the physical distancing measures. The unexpected threshold phenomenon revealed by our analysis highlights our unfamiliarity with the mental health consequences of those measures and the pressing need for more quantitative studies on the interplay between mood, behavior, and productivity.

Acknowledgments This research was supported in part by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) under Grant No. 2020/03041-3 and by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) under Grant No. 305058/2017-7.

References

1. D.L. Phillips, Social participation and happiness. Am. J. Sociol. **72**, 479–488 (1967)
2. J.T. Cacioppo, L.C. Hawkley, Perceived social isolation and cognition. Trends Cogn. Sci. **13**, 447–454 (2009)
3. A. Oswald, E. Proto, D. Sgroi, Happiness and productivity. J. Labor Econ. **33**, 789–822 (2015)
4. A. Pentland, How social networks network best. Harv. Bus. Rev. **87**, 37 (2009)

5. T. Cornelissen, C. Dustmann, U. Schönberg, Peer effects in the workplace. Am. Econ. Rev. **107**, 425–456 (2017)
6. E. Kandel, E.P. Lazear, Peer pressure and partnerships. J. Polit. Econ. **100**, 801–817 (1992)
7. O. Bandiera, I. Barankay, I. Rasul, Social incentives in the workplace. Rev. Econ. Stud. **77**, 417–458 (2010)
8. M.O. Feldman, D.B. Audretsch, Innovation in cities: science based diversity, specialization and localized competition. Eur. Econ. Rev. **43**, 409–429 (1999)
9. L.M.A. Bettencourt, J. Lobo, D. Helbing, C. Kühnert, G.B. West, Growth, innovation, scaling, and the pace of life in cities. Proc. Natl. Acad. Sci. USA **104**, 7301–7306 (2007)
10. B.G. Smith, S.B. Smith, D. Knighton, Social media dialogues in a crisis: a mixed-methods approach to identifying publics on social media. Public Relat. Rev. **44**, 562–573 (2018)
11. S. Banskota, M. Healy, E.M. Goldberg, 15 smartphone apps for older adults to use while in isolation during the COVID-19 pandemic. West. J. Emerg. Med. **21**, 514–525 (2020)
12. L. Liang, H. Ren, R. Cao, Y. Hu, Z. Qin, L. Chuanen, M. Songli, The effect of COVID-19 on youth mental health. Psychiatr. Q. **91**, 841–852 (2020)
13. E.D. Miller, Cyberloneliness: the curse of the cursor? in *Narratives of Loneliness: Multidisciplinary Perspectives from the 21st Century*, ed. by O. Sagan, E.D. Miller (Routledge, London, 2018), pp. 56–65
14. J.N. Bailenson, Nonverbal overload: a theoretical argument for the causes of zoom fatigue. Technol. Mind Behav. **2** (2021). https://doi.org/10.1037/tmb0000030
15. S.M. Moorman, Dyadic perspectives on marital quality and loneliness in later life. J. Soc. Personal Relat. **33**, 600–618 (2016)
16. I. Asimov, *Robot Trilogy: The Caves of Steel, The Naked Sun, The Robots of Dawn* (Del Rey, New York, 1988)
17. N. Bellomo, R. Bingham, M.A.J. Chaplain, G. Dosi, G. Forni, D.A. Knopoff, J. Lowengrub, R. Twarock, M.E. Virgillito, A multiscale model of virus pandemic: heterogeneous interactive entities in a globally connected world. Math. Models Methods Appl. Sci. **30**, 1591–1651 (2020)
18. E.D. Miller, Loneliness in the Era of COVID-19. Front. Psychol. **11**, 2219 (2020)
19. P. Hardy, L.S. Marcolino, J.F. Fontanari, The paradox of productivity during quarantine: an agent-based simulation. Eur. Phys. J. B **94**, 40 (2021)
20. J.F. Fontanari, A stochastic model for the influence of social distancing on loneliness. Phys. A **584**, 126367 (2021)
21. A.M. Isen, K.A. Daubman, G.P. Nowicki, Positive affect facilitates creative problem solving. J. Pers. Soc. Psychol. **52**, 1122–1131 (1987)
22. H.C. Ellis, P.W. Ashbrook, Resource allocation model of the effects of depressed mood states on memory, in *Affect, Cognition, and Social Behavior*, ed. by K. Fiedler, J. Forgas (Hogrefe, Göttingen, 1988), pp. 35–42
23. C.W. McIntyre, D. Watson, L.A. Clark, S.A. Cross, The effect of induced social interaction on positive and negative affect. Bull. Psychon. Soc. **29**, 67–70 (1991)
24. S. von Stumm, Feeling low, thinking slow? Associations between situational cues, mood and cognitive function. Cogn. Emot. **32**, 1545–1558 (2018)
25. G. Johns, A.M. Saks, *Organizational Behaviour: Understanding and Managing Life at Work* (Pearson/Prentice Hall, Toronto, 2005)
26. K. Huang, *Statistical Mechanics* (Wiley, New York, 1963)
27. F.B. Alberti, *A Biography of Loneliness: The History of an Emotion* (Oxford University Press, New York, 2019)
28. P. Mac Carrona, K. Kaskib, R. Dunbar, Calling Dunbar's numbers. Soc. Netw. **47**, 151–155 (2016)
29. U.K. Age, Jo Cox commission final report (2017). https://www.ageuk.org.uk/globalassets/age-uk/documents/reports-and-publications/reports-and-briefings/active-communities/rb_dec17_jocox_commission_finalreport.pdf
30. J.D. Murray, *Mathematical Biology I: An Introduction* (Springer, New York, 1993)

31. N. Leigh-Hunt, D. Bagguley, K. Bash, V. Turner, S. Turnbull, N. Valtorta, W. Caan, An overview of systematic reviews on the public health consequences of social isolation and loneliness. Public Health **152**, 157–171 (2017)
32. N. Bloom, J. Liang, J. Roberts, Z.J. Ying, Does working from home work? Evidence from a Chinese experiment. Q. J. Econ. **130**, 165–218 (2015)
33. P. Courtet, E. Olié, C. Debien, G. Vaiva, Keep socially (but not physically) connected and carry on: preventing suicide in the age of COVID-19. J. Clin. Psychiatry **81**, 20com13370 (2020)
34. T. Hogg, B.A. Huberman, Better than the best: the power of cooperation, in *1992 Lectures in Complex Systems*, ed. by L. Nadel, D. Stein (Addison-Wesley, Reading, 1993), pp. 165–184
35. S.M. Reia, P.F. Gomes, J.F. Fontanari, Policies for allocation of information in task-oriented groups: elitism and egalitarianism outperform welfarism. Eur. Phys. J. B **92**, 205 (2019)
36. S.M. Reia, A.C. Amado, J.F. Fontanari, Agent-based models of collective intelligence. Phys. Life Rev. **31**, 320–331 (2019)
37. J.F. Fontanari, Imitative learning as a connector of collective brains. PLoS ONE **9**, e110517 (2014)

Transmission Dynamics and Quarantine Control of COVID-19 in Cluster Community

Qingwu Gao, Jun Zhuang, Ting Wu, and Houcai Shen

1 Introduction

Coronavirus Disease 2019 (COVID-19) is a recently described zoonotic disease with a high infection rate which has spread rapidly and widely all over the world in the past two years [9, 22]. As of September 7, 2021, the cumulative number of infectious cases reported globally is now just over 220 million and the cumulative number of deaths is over 4.5 million [4, 5]. The causative agent of COVID-19 is identified as a novel coronavirus, which so far has four variants of concern: Alpha, Beta, Gamma, and Delta, and then the severity of COVID-19 has far exceeded that of SARS in 2003 [13]. The COVID-19 outbreak has attracted considerable global attention. On March 11, 2020, the World Health Organization (WHO) officially announced that this epidemic is a global pandemic, and since then has issued a

Q. Gao (✉)
School of Statistics and Data Science, Nanjing Audit University, Nanjing, China

Department of Industrial and Systems Engineering, University at Buffalo, Buffalo, NY, USA

School of Management Science and Engineering, Nanjing University, Nanjing, China
e-mail: qwgao@aliyun.com; qwgao@nau.edu.cn

J. Zhuang
Department of Industrial and Systems Engineering, University at Buffalo, Buffalo, NY, USA
e-mail: jzhuang@buffalo.edu

T. Wu (✉)
Department of Mathematics, Nanjing University, Nanjing, China
e-mail: tingwu@nju.edu.cn

H. Shen
School of Management Science and Engineering, Nanjing University, Nanjing, China
e-mail: hcshen@nju.edu.cn

N. Bellomo, M. A. J. Chaplain (eds.), *Predicting Pandemics in a Globally Connected World, Volume 1*, Modeling and Simulation in Science, Engineering and Technology, https://doi.org/10.1007/978-3-030-96562-4_8

series of interim guidelines to all countries on how to prepare for this emergency, for example, how to monitor potential infected individuals, collect and test samples, effectively communicate with the public on the COVID-19 [20]. And in China, the government has implemented some unprecedented and extreme public-health intervention strategies, including travelling and gathering restrictions, intensive contact tracing followed by quarantine of individuals potentially exposed to the disease, isolation of infected and symptomatic individuals, and free vaccinations against this disease [18].

The novel coronavirus, which is the causative agent of COVID-19, is mainly transmitted by close aerosol and close contact and has a significant aggregation phenomenon in families and hospitals which are of a feature that they are all clustered-populated, relatively confined, and isolated from society, with the name "Cluster Community" [3]. The extent of the epidemic of COVID-19 has been worsened by the occurrence of large clusters of infection linked to single individuals and/or spatial locations. As a typical example, we take the Cruise ship "Diamond Princess," one of the most luxurious cruise ships in the world, which has recently received extensive attention due to its growing epidemic of COVID-19. After the pneumonia-related symptoms appearing in the first infectious case on January 23, 2020, the novel coronavirus has been spreading rapidly on the Cruise ship so that 712 individuals, or about 20% of the total, have been confirmed finally to be infected by the COVID-19 [4, 11].

With the resuming of social and economic activities while continuing COVID-19 transmission throughout the world, large masses of people are expected to pour into their workplaces, schools, and other cluster communities, which can increase the potential for an infectious agent to rebound substantially. Then, how to effectively prevent, detect, and defend against the COVID-19 infection in the cluster community is an extremely urgent and vital duty for the government officials. Public-health efforts have been taken, for example, on February 5, 2020, the "Diamond Princess" cruise ship was performed quarantine, and 10 infectious cases were sent to hospitals for treatment; On February 13, an intervention measure that passengers can choose to disembark for quarantine was issued, and on February 15, the authorities announced the disembarkation instructions [10]. Acknowledging that the quarantine intervention is so expensive and resource-intensive that it could interrupt social-economic activities to some extent, quantifying the effectiveness of quarantine is of crucial importance for the "Diamond Princess" and other similar cluster communities to prevent and control the outbreak of COVID-19.

As human, material, and financial resources are limited, governments and their officials would like to identify the most appropriate, evidence-based interventions and mitigation measures and to determine the intensity of such programs. For a cluster community like the Cruise ship "Diamond Princess," the quarantine intervention could be affected by the following three key factors. The first one is the current spreading scope of COVID-19. If the novel coronavirus that is the causative agent of COVID-19 is just beginning to circulate in a small scope, less intervention effort is needed. The second factor is the popularity of COVID-19. For super transmitters with the greatest capacity to infect counter-parties, the transmission

could be surprisingly fast, and more quarantine effort should be made. The third factor is the effectiveness of different quarantine strategies. As the effectiveness of these interventions could be different for COVID-19 with different scope and popularity, it is imperative to investigate the efficiency of different quarantine strategies.

Considering that an effective quarantine intervention may be a potential tool to mitigate the magnitude and the speed of the COVID-19 outbreak, it is a study of great importance and considerable interest not only to accurately predict the tendency of COVID-19 transmission in the host population but also to effectively fight against this prevalence by quarantine measure in a cluster community. In the past decades, researchers have proposed mathematical models in an epidemiology sense to describe the transmission dynamics of infectious disease and to discuss the impacts of public-health interventions on reducing the size of the transmission [8, 12]. As a highly contagious infectious disease, COVID-19 has been investigated in an extensive literature after its outbreak throughout the world. For example, Yang et al. [21] integrated real epidemiological data into the SEIR model to derive the epidemic curve and used an AI approach to predict the epidemic and suggested that a five-day delay in implementation would have increased epidemic size in mainland China threefold. Kraemer et al. [7] used real-time mobility data from Wuhan and detailed case data including travel history to elucidate the role of case importation in transmission in cities across China and to ascertain the impact of control measures, which shows that the control measures implemented in China substantially mitigated the spread of COVID-19. Tian et al. [19] investigated the spread and control of COVID-19 by using a data set that included case reports, human movement, and public-health interventions, which indicates that Wuhan shutdown was associated with the delayed arrival of COVID-19 in other cities by 2.91 days, and that cities that implemented control measures reported fewer cases on average (13.0) compared with cities that started control later (20.6). Bellomo et al. [1] proposed a multi-scale model of COVID-19 pandemic to consider the interaction of different spatial scales, from the small scale of the virus itself and cells to the large scale of individuals and further up to the collective behavior of populations, which develops an interdisciplinary vision due to the contributions of epidemiologists, immunologists, economists, and mathematical modelers.

We notice that the references mentioned above characterize the dynamics of the COVID-19 spreading and discuss the impacts of control measures from a macro point of view. To the best of our knowledge, there is no existing literature to consider the interplay between transmission dynamics and quarantine intervention in a cluster community with crowded population and confined spaces. Hence in this study, we apply a non-linear ordinary differential equations (ODEs) system to describe the transmission dynamics of COVID-19 and to capture the impact of quarantine intervention, where the key characteristics of the spreading and control of COVID-19 are parameterized. With the use of detailed epidemiologic data from the Cruise ship "Diamond Princess," we design a Transmission-Quarantine workflow to determine the case-specific parameters and validate the proposed model by comparing the simulated curve with the real data. Firstly, we propose a general

SEIR-type epidemiological model to investigate the transmission dynamics of COVID-19 without quarantine intervention and present the analytic and simulation results for the epidemiological parameters such as the basic reproduction number, the maximal scale of infectious cases, the instant number of recovered cases, the popularity level, and the final scope of the epidemic of COVID-19. Secondly, we adopt the proposed Transmission-Quarantine interplay mechanism to predict the varying trend of COVID-19 with quarantine intervention and compare the transmission dynamics with and without quarantine to illustrate the effectiveness of the quarantine measure. Finally, we conduct sensitivity analysis to simulate the impacts of different parameters and different quarantine measures and identify the optimal quarantine strategy that will be used by the decision makers to achieve the maximal protection of population with the minimal interruption of economic and social development.

The rest of this chapter is organized as follows. In Sect. 2, we formulate the mathematical models for the COVID-19 outbreak in a cluster community with and without quarantine intervention. The main results of this chapter, i.e., analytic results, numerical results and case analyses, are presented in Sects. 3 and 4, respectively, followed by discussion and concluding remarks in Sect. 5. The corresponding proofs of this study are available in the Appendix.

2 Mathematical Modeling

According to the clinical progression of COVID-19, epidemiological status of the individuals, and the performed quarantine measure, the life-cycle of the COVID-19 epidemic in a cluster community could be divided into two stages. The first stage is the epidemic emerging stage, which refers to the time period between the point when the first infectious case emerges and the point when the quarantine invention is performed. The second stage is the epidemic quarantine stage, which refers to the time period between the quarantine point and the epidemic ending point, namely that the point when all the infected cases are recovered. The quarantine measure implemented in the second stage can reduce the risk of exposure of individuals to the novel coronavirus, and then the magnitude and the speed of COVID-19 outbreak can decrease. We define the end point of the life-cycle as the point of time when the number of the infectious cases becomes under one. The life-cycle of the COVID-19 epidemic is shown in Fig. 1.

In the emerging stage, the whole population in a cluster community is stratified as susceptible (S), latent or exposed (E), infectious (I), and recovered (R) groups; and further in the quarantine stage, the population also consists of the quarantined susceptible (S_Q) and quarantined exposed (E_Q) groups. In this chapter, we assume that the population of the "Diamond Princess" cruise ship is relative homogeneous mixing [6], and thus the coronavirus transmission process could be modeled at a macro level. The definitions of all the above groups are summarized in Table 1.

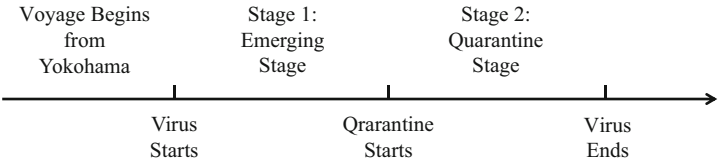
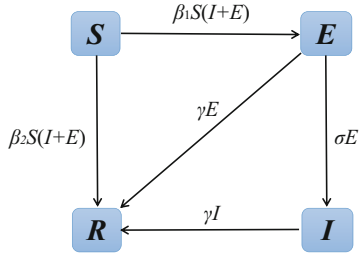


Fig. 1 Transmission process of COVID-19 in a community

Table 1 Definitions of the groups

Group	Definition
Susceptible (S)	People who have not been infected by COVID-9
Latent/Exposed (E)	People who have been infectious but not yet symptomatic in the latent time
Infectious (I)	People who are infectious confirmed by their symptoms or by clinical diagnosis
Quarantined Susceptible (S_Q)	Quarantined people who are not infected
Quarantined Exposed (E_Q)	Quarantined and infectious cases without symptoms
Recovered (R)	People who have recovered from COVID-19

Fig. 2 Transition diagram in Stage 1



2.1 Stage 1: SEIR-Type Model Without Quarantine

In this section, we propose a general SEIR-type epidemic model to study the transmission dynamics of COVID-19 without quarantine effort in a cluster community. At the initial time point ($t = 0$), the population is all susceptible, except the initial one infectious case. The proportions of the population in S, E, I, and R groups at time t are $S(t)$, $E(t)$, $I(t)$, and $R(t)$, respectively, with the normalization condition: $S(t) + E(t) + I(t) + R(t) = 1$. Note that the entire population N of the cluster community remains constant with time since this community is a closed and isolated unit from society during the COVID-19 epidemic. The block diagram of SEIR-type model with modeling parameters is shown in Fig. 2. Table 2 shows notation and parameters used in this model, and Table 3 shows a summary of the group transitions and transition rates in Stage 1.

Combining Fig. 2, Tables 2 and 3, the transitions of population in Stage 1 are detailed as follows:

Table 2 Notation and parameters in SEIR-type model

Notation	Explanation
N	Whole population in a cluster community
S(t)	Fraction of susceptible individuals at time t
E(t)	Fraction of latent individuals at time t
I(t)	Fraction of infectious individuals at time t
R(t)	Fraction of recovered individuals at time t
β_1	Infection rate of the susceptible by COVID-19
β_2	Rate of the susceptible contacted but uninfected by COVID-19
σ	Transition rate from E to I
γ	Transition rate from E/I to R

Table 3 Transition events and transition rates in Stage 1

Events	Rates of transition
S→E	$\beta_1 \mathrm{S}(t)(\mathrm{I}(t) + \mathrm{E}(t))$
S→R	$\beta_1 \mathrm{S}(t)(\mathrm{I}(t) + \mathrm{E}(t))$
E→I	$\sigma \mathrm{E}(t)$
E→R	$\gamma \mathrm{E}(t)$
I→R	$\gamma \mathrm{I}(t)$

1. As a latent case of COVID-19 with no symptoms is of actively infectious capacity, the infectious cases mentioned in this study include those who are the latent cases and the confirmed infectious cases. When an infectious case contacts a susceptible individual, the susceptible one may become a latent one with the rate β_1. Thus in the equation system, the proportion of individuals who transit from group S to group E per time unit is $\beta_1 S(t)(I(t) + E(t))$.
2. When an infectious case contacts a susceptible individual, the susceptible one may not be infected by the new coronavirus and become a recovered one with the rate β_2. The proportion of individuals who transit from group S to group R per time unit is $\beta_2 S(t)(I(t) + E(t))$.
3. A latent case transits into group I with rate σ when the latent one has symptoms of COVID-19 or is detected by clinical diagnosis, where $1/\sigma$ is the mean latent period.
4. Using some sufficient treatment interventions, a latent case that is infectious but asymptomatic one, or a confirmed infectious case becomes recovered at rate γ, where $1/\gamma$ is the mean cure period through treatment interventions.

In the emerging stage before quarantine intervention, the transmission dynamics of COVID-19 spreading in a cluster community is governed by the system of nonlinear ODEs as follows:

$$\begin{cases} S'(t) = -(\beta_1 + \beta_2)S(t)(I(t) + E(t)), \\ E'(t) = \beta_1 S(t)(I(t) + E(t)) - (\sigma + \gamma)E(t), \\ I'(t) = \sigma E(t) - \gamma I(t), \\ R'(t) = \beta_2 S(t)(I(t) + E(t)) + \gamma(E(t) + I(t)). \end{cases} \tag{1}$$

Then in the first stage without quarantine control, we are interested in studying the following questions:

1. Predicting the final scope of COVID-19 to access the cumulative close contacts of infectious cases. In other words, without being intervened by quarantine measure, how many people in a cluster community will be finally contacted by the novel coronavirus/COVID-19? Moreover, we study how the rates of transition among the four groups influence the final scope.
2. Predicting the maximal scale of the infectious cases, including the latent and confirmed infectious cases. In other words, we aim to predict when would the COVID-19 infection process arrive its peak, and what will be the maximal number of infectious cases that exist? In this way, the obtained results can help decision makers to evaluate the population of the asymptomatic and infectious cases.
3. Calculating the population of recovered cases at any time point, which can help decision makers to estimate how many people have already contacted the infectious cases but not necessarily been infected by the new coronavirus.
4. Estimating the popularity level of COVID-19 epidemic. Once an infectious case is detected, our model can estimate how fast the new coronavirus is transmitting in a cluster community.

2.2 Stage 2: Transmission-Quarantine (TQ) Model

In this section, we consider the second stage of the COVID-19 epidemic with quarantine invention. More specifically, a proportion, $q, 0 \leq q \leq 1$, of susceptible individuals is quarantined, and then the transmission of COVID-19 is impacted by the quarantine intervention. Figure 3 indicates the diagram of the Transmission-Quarantine (TQ) epidemic model in Stage 2, involving group transition and modeling parameters, Table 4 presents the modeling parameters in TQ model, and Table 5 presents the group transitions and rates of transition in Stage 2.

Following the classical virus spreading model [6], the proposed TQ model starts with a relative homogeneous mixing population in a cluster community. For a specific virus/disease, the whole population in the community is assumed to be fixed, and the infectious duration and infectious probability of the population are homogeneous. Thus, we consider the novel coronavirus/COVID-19 transmission at a macro level. Denote the proportions of the populations in the groups in Fig. 3 by

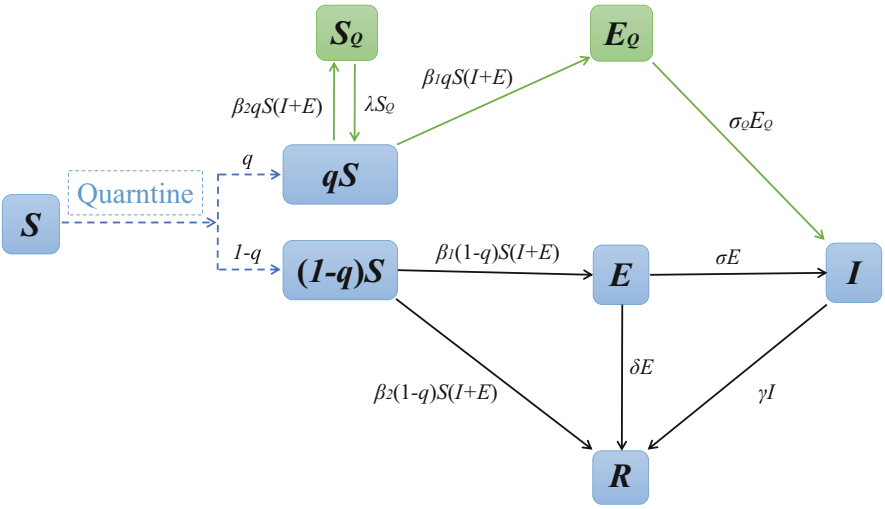


Fig. 3 Transition diagram in Stage 2

Table 4 Parameters of TQ model except those in Stage 1

Parameters	Explanation
q	The quarantine proportion
λ	Transition rate from S_Q to S
σ_Q	Transition rate from E_Q to I

Table 5 Transition events and transition rates in Stage 2

Events including quarantine intervention	Rates of transition
$S \to E$	$(1-q)\beta_1 S(t)(I(t)+E(t))$
$S \to R$	$(1-q)\beta_2 S(t)(I(t)+E(t))$
$S \to S_Q$	$q\beta_2 S(t)(I(t)+E(t))$
$S \to E_Q$	$q\beta_1 S(t)(I(t)+E(t))$
$S_Q \to S$	$\lambda S_Q(t)$
$E \to I$	$\sigma E(t)$
$E \to R$	$\gamma E(t)$
$E_Q \to I$	$\sigma_Q E_Q(t)$
$I \to R$	$\gamma I(t)$

$S(t)$, $E(t)$, $I(t)$, $S_Q(t)$, $E_Q(t)$, and $R(t)$, respectively, satisfying the normalization condition: $S(t)+E(t)+I(t)+S_Q(t)+E_Q(t)+R(t)=1$, which means that the cluster community has a fixed population because of its closed and isolated feature during the COVID-19 epidemic.

From Fig. 3, Tables 4 and 5, we present the group transitions in TQ model of Stage 2 as follows:

1. With contact tracing, the quarantine intervention is implemented to reduce the risk of exposure of individuals to the novel coronavirus, a proportion q, $0 \leq$

$q \leq 1$, individuals exposed to the virus is quarantined. Then the proportion of quarantined ones is $qS(t)$ in all. When $q = 0$, the quarantine intervention is not implemented or has no effect at all; when $q = 1$, the quarantine intervention is perfectly effective.

2. The quarantined individuals can either move to the group E_Q or S_Q, depending on whether they are effectively infected or not by COVID-19. Specifically, when a quarantined one contacts an infectious case, she/he is effectively infected by the disease and then becomes a latent one with the rate β_1. Thus the proportion of those who transit to quarantined exposed group E_Q per time unit is $q\beta_1 S(t)(I(t) + E(t))$. Otherwise, the quarantined one is not infected and moves to the quarantined susceptible group S_Q at rate β_2. Thus the proportion of those who transit to the group S_Q per time unit is $q\beta_2 S(t)(I(t) + E(t))$.
3. Considering that the great cost of human and material resources for quarantine intervention, the quarantined susceptible individuals will be released from quarantine control and transited to the susceptible group with rate λ, where $1/\lambda$ is the mean quarantine period.
4. Besides the susceptible individuals quarantined, the other proportion, $1 - q$, of susceptible ones missed from the contact tracing and move to the exposed group E once effectively infected or move to the recovered group R otherwise. Specifically, when the infectious case, either the latent one or the confirmed infectious one, contacts a susceptible individual who is not quarantined, the susceptible one may become a latent one with the rate β_1. Thus in the equation system, the proportion of those who transit from the susceptible group S to the exposed group E per time unit is $(1 - q)\beta_1 S(t)(I(t) + E(t))$. Otherwise, the susceptible one may not be infected by the coronavirus, and become a recovered one with the rate β_2. The proportion of those who transit from group S to group R per time unit is $(1 - q)\beta_2 S(t)(I(t) + E(t))$.
5. As time goes by, a latent case in groups E and E_Q transits into infectious group I with rates σ and σ_Q, respectively, when the latent one is infectious confirmed by COVID-19 symptoms or detected by clinical diagnose.
6. After some sufficient treatment interventions, an infectious case, including a latent one or a confirmed one, becomes recovered at rate γ.

According to the statements mentioned above, the TQ model with quarantine intervention can be formulated by the following non-linear ODEs system as:

$$\begin{cases} S'(t) = -(\beta_1 + \beta_2)S(t)(I(t) + E(t)) + \lambda S_Q(t), \\ E'(t) = (1 - q)\beta_1 S(t)(I(t) + E(t)) - (\sigma + \gamma)E(t), \\ I'(t) = \sigma E(t) + \sigma_Q E_Q(t) - \gamma I(t), \\ S'_Q(t) = q\beta_2 S(t)(I(t) + E(t)) - \lambda S_Q(t), \\ E'_Q(t) = q\beta_1 S(t)(I(t) + E(t)) - \sigma_Q E_Q(t) \\ R'(t) = (1 - q)\beta_2 S(t)(I(t) + E(t)) + \gamma(E(t) + I(t)). \end{cases} \tag{2}$$

In Sect. 4, we will use Runge-Kutta methods [2] to solve the differential equation system (2), and analyze the impacts of different modeling parameters (e.g., transition rates) on the transmission process of COVID-19.

3 Analytic Results and Case Study for Emerging Stage

3.1 Analytic Results

Before the quarantine intervention, the novel coronavirus spreads fast and infects almost individuals in a crowded and closed community. Hence, we firstly investigate the expected final scope of COVID-19 caused by the novel coronavirus. In other words, without being intervened by quarantine, how many people will contact the infectious cases in the end? Let $\hat{R}$ be the final scale of the epidemic of COVID-19, i.e., the fraction of recovered group R in the entire population at the end, $\hat{R} =: R(\infty) \in (0, 1)$, which can be used to estimate the number of close contacts tracked cumulatively of COVID-19. By investigating the ODEs system (1), we draw the two conclusions below. Because of the technical complexity, the formal proofs are delayed to Appendix.

Theorem 1 *Without quarantine intervention, $\hat{R}$ satisfies the following equation:*

$$\hat{R} = 1 - \exp\left\{-\frac{\beta_1}{\gamma}\hat{R}\right\}. \tag{3}$$

If $\frac{\beta_1}{\gamma} > 1$, Eq. (3) has two solutions: zero and non-trivial solution $\hat{R}$, $0 < \hat{R} < 1$.

Theorem 2 *Given a fixed γ, the non-trivial solution $\hat{R}$ increases as β_1 increases. Similarly, given a fixed β_1, $\hat{R}$ decreases as γ increases.*

Remark 1 The above two theorems provide some insights to the COVID-19 transmission in the first emerging stage. Theorem 1 gives the implicit equation of the final scope $\hat{R}$, which indicates that the final size of the COVID-19 transmission is only dependent on two factors: β_1 and γ. The value of solution $\hat{R}$ and its relation with β_1 and γ are shown in Fig. 4a. In summary, when $\frac{\beta_1}{\gamma} > 1$, the final scale $\hat{R}$ can be non-zero, which means the new coronavirus would circulate among the population; when $\frac{\beta_1}{\gamma} < 1$, this final scale is zero, which means the coronavirus would not start circulating. The 3D plot and the contour plot in Fig. 4b, c further illustrate how the final scale $\hat{R}$ vary over β_1 and γ.

It is shown from Fig. 4a that the parameter $\frac{\beta_1}{\gamma}$ plays a key role to determine the final scope of the COVID-19 outbreak, which implies a theorem as follows:

Theorem 3 *Without quarantine intervention, the outbreak of COVID-19 will remain under control when $\frac{\beta_1}{\gamma} < 1$ and uncontrollable when $\frac{\beta_1}{\gamma} > 1$.*

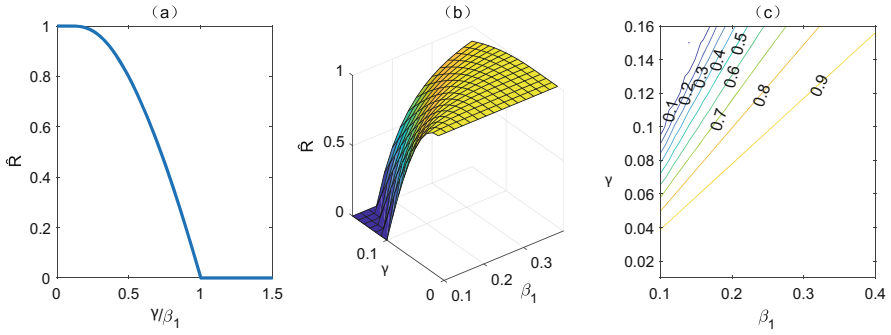


Fig. 4 Analysis of final scale $\hat{R}$ of COVID-19

Remark 2 The parameter $\frac{\beta_1}{\gamma}$ can be called the transmission control coefficient of COVID-19. This coefficient is an important and useful parameter in sense that it provides information about the COVID-19 transmission process, and it can be used to predict whether the disease will propagate among the population or not. The coefficient corresponds to the threshold value, known as the basic reproduction number [8, 12] that can be helpful for determining the potential and severity of an outbreak, and providing critical information for identifying the type of disease interventions and intensity. Theorem 3 suggests when $\frac{\beta_1}{\gamma} > 1$, the COVID-19 invades the population; when $\frac{\beta_1}{\gamma} < 1$, the COVID-19 will gradually disappear, and then quarantine intervention may not be needed.

Subsequently, we devote ourselves to investigate the maximum scale of the infectious cases, which include the latent cases and the confirmed cases. We define $I_{\max}(t)$ as the maximum scale of the infectious cases, which satisfies the following theorem.

Theorem 4 *Without quarantine intervention, when* $S(t) = \frac{\gamma}{\beta_1}$, *the number of infectious cases reaches its maximum scale as:*

$$I_{\max}(t) = \frac{\gamma}{\beta_1 + \beta_2}\left(\ln\frac{\gamma}{\beta_1} - 1\right) + \frac{\beta_1}{\beta_1 + \beta_2}.$$

From Theorem 4, it follows that at the point when the scale of the infectious cases (i.e., $E(t) + I(t)$) reaches its maximum value, the proportion of the population in recovered group R is $R(t) = 1 - \frac{\gamma}{\beta_1} - I_{\max}(t)$. Finally, we discuss the impact of the parameter $\frac{\gamma}{\beta_1}$ on the dynamic behavior of recovered cases with respect to susceptible individuals.

Theorem 5 *Without quarantine intervention, when* $S(t) \geq \frac{\gamma}{\beta_1}$, *the scale of recovered cases,* $R(t)$, *increases slowly as* $S(t)$ *decreases; when* $S(t) < \frac{\gamma}{\beta_1}$, $R(t)$ *increases sharply as* $S(t)$ *decreases.*

Note that the parameter $\frac{\gamma}{\beta_1}$ is a sharp threshold value at which the transition from case that $R(t)$ increases slowly to that $R(t)$ increases more quickly occurs, as $S(t)$ decreases. The result shows the threshold value could play a critical role to control the epidemic of COVID-19. Precisely, the decision makers should perform their corresponding public-health interventions before the scale of susceptible individuals reaches $\frac{\gamma}{\beta_1}$; otherwise, if the scale is smaller than $\frac{\gamma}{\beta_1}$, the novel coronavirus will spreading sharply throughout the whole population, and the COVID-19 outbreak is totally out of control.

The third and fourth questions stated in Sect. 2.1 are analyzing the current scope and popularity level of the COVID-19 epidemic. In other words, how many people have been contacted by the infectious cases at any time, how many people are infectious but pre-symptomatic in incubation period, and how fast the novel coronavirus is transmitting in a cluster community? In the real world, government officials and other decision makers can only know the number of the confirmed infectious cases of COVID-19, including confirmed cases with symptoms, clinically diagnosed cases, and even suspected cases. However, the majority of the population who have been contacted by the infectious cases may not have any observable spreading behavior. Moreover, it is difficult to directly observe how fast the novel coronavirus will keep spreading. So in the following subsection, we demonstrate our solution to these questions by numerically solving the ODEs system by using real world case study. We design an implementation work-flow to derive the optimal solution. The case study validates our model with real life data, and our model could predict the future trend of the COVID-19 spreading.

3.2 A Real World Case Study for Stage 1

The Cruise ship "Diamond Princess" is one of the world's top luxury cruises with the largest volume and the most complete facilities, just like a five-star hotel on the sea. Recently, however, the Cruise ship has attracted global attention due to its increasingly serious COVID-19 outbreak. The "Diamond Princess" with 2666 passengers and 1045 crew onboard was departed from Yokohama, Japan, on January 20, 2020. And on February 1, a passenger who got off the ship on January 25 was hospitalized with fever and diagnosed with COVID-19. This is the first confirmed infectious case in the cruise ship, and the passenger developed pneumonia-related symptoms on January 23. The ship abandoned its original journey and made an emergency berth near Yokohama on February 3, and quarantine intervention was performed for 14 days by Japanese government on February 5. During the quarantine, all the undiagnosed passengers and crew were not allowed to disembark, and only stay in their room to prevent the outbreak of COVID-19. By the end of quarantine period on February 19, a total of 542 people were confirmed infectious, and the passengers officially began to disembark. As of March 15, 712 cases, or about 20% of the total, have been confirmed finally to be infected by COVID-

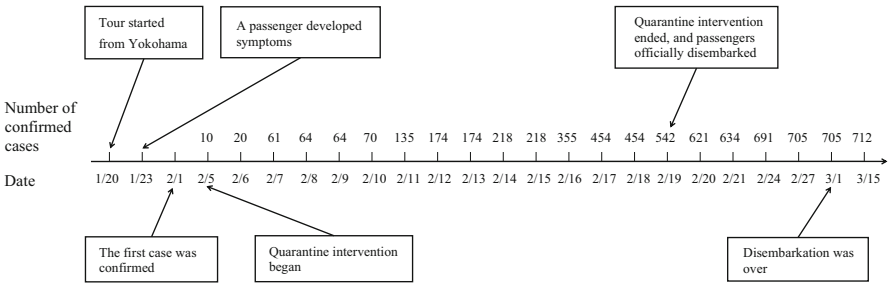


Fig. 5 Timeline and events of COVID-19 in "Diamond Princess"

19. The timeline and related events of the COVID-19 epidemic in the "Diamond Princess" cruise ship are shown in Fig. 5.

The data employed in the chapter are official data provided by WHO [4], Coronavirus Resource Center of John Hopkins University [5], Ministry of Health, Labor and Welfare of Japan [11], and the official website of Diamond Princess [15]. Moreover, we manually collect the time stamps and the counts of infectious cases from January 20 to February 22, which is 3 days after the end of quarantine period on February 19 since the report date is approximately 3 days after the onset date [11]. One day is set as the unit of time in this chapter. The data collected here can provide the variation trend of the infectious cases, which is used to determine the parameter values in the proposed models, to simulate the whole spreading process of the novel coronavirus, and further to estimate the scope of influence of the COVID-19 epidemic. It should be mentioned that in our models, the main variables $I(t)$ and $R(t)$ are the proportions of the infectious and recovered cases in each time step. The two variables are measured by the population of the infectious and recovered groups within each time interval.

In the following, we will design a work-flow to determine the case-specific parameters of the epidemic emerging stage by our real life data. The work-flow of numerical simulation is shown in Fig. 6. Firstly, a set of initial parameters are put into system (1). Secondly, solving the ODEs system by the initial parameters, we obtain the simulated variation trend of $I(t)$. Thirdly, we calculate the error term $|I(t) - I^{\text{real}}(t)|$. Finally, by iteratively updating the parameters, we minimize the error term and obtain the optimal set of parameters that best fits the COVID-19 transmission process.

Now we use the analysis below to mathematically identify the real start time of the COVID-19 outbreak in the "Diamond Princess" cruise ship. Consider that there is only one infectious case at the very beginning of the voyage from Yokohama, we know that when $t = 0$, $S(0) = \frac{N-1}{N}$, $E(0) = 0$, $I(0) = \frac{1}{N}$, and $R(0) = 0$, where N is the whole population in the cruise ship. At the early stage of the coronavirus transmission, the numbers of latent, infectious, and recovered cases are negligible relatively to the population N. Thus, the proportion of susceptible individuals in the early stage is approximately invariant, i.e., $S(t) \approx 1$. By summing

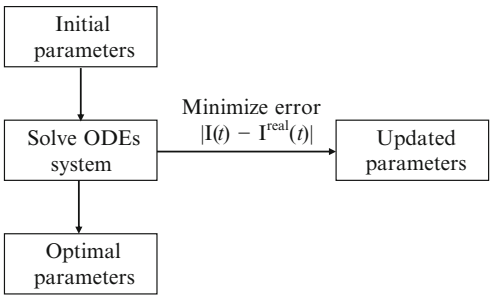


Fig. 6 Work-flow of numerical simulation

the second and third equations of system (1) and taking the integral, we have that $I(t) = \frac{1}{N}\exp\{(\beta_1 - \gamma)(t + t_0)\}$, where the notation t_0 means that the date earlier t_0 days than January 20 is the real start date of the COVID-19 epidemic in the cruise ship. The number of infectious cases is approximately $\exp\{(\beta_1 - \gamma)(t + t_0)\}$, which implies that when $\beta_1 < \gamma$, the number of infectious cases is less than 1, and then the outbreak of COVID-19 will disappear; and when $\beta_1 > \gamma$, the infectious cases is persistent, and then the outbreak is uncontrollable. This conclusion is consistent with Theorems 1 and 3.

We adopted Matlab as a tool to apply the curve fitting of the real data to minimize $|I(t) - I^{\text{real}}(t)|$, which is shown in Fig. 7. The dot is the real data point of infectious cases in each time point, and the line is the fitted function curve which is $\exp\{(\beta_1 - \gamma)(t + t_0)\} \approx \exp\{0.6188(t - 13)\}$. The result shows a good match with the real data with a goodness of fit R-square = 0.9866. Hence, the estimated parameters satisfy that $\beta_1 - \gamma = 0.6188$, and $t_0 = -13$ indicates that the date later 13 days than January 20 is the real start date, namely February 1, of the epidemic of COVID-19. In terms of the real data, the first infectious passenger in the "Diamond Princess" was confirmed on February 1. However, there is no new infectious case confirmed up to February 5, which is because the incubation period of the new coronavirus is about 14 days. In fact, from February 5 this epidemic is gradually blooming. The simulation result is in line with the real data from the Cruise ship "Diamond Princess."

The total number of all passengers and crew in the Cruise ship "Diamond Princess" is $N = 3711$. From Lauer et al. [7], the estimated median incubation period of COVID-19 was 5.1 days, then the transition rate from groups E to I is $\sigma = 1/5.1$. According to the press conference of National Health Commission (NHC) of China on February 4, 2020 [10], the average cure period is about 9.5 days. Then the transition rates from E to R and from I to R are both $\gamma = 1/9.5$. As $\beta_1 - \gamma = 0.6188$, the infected rate of susceptible individuals by the novel coronavirus is $\beta_1 = 0.7241$. Hence, the basic reproduction number of the COVID-19 epidemic in the "Diamond Princess" is estimated as $R_0 = 6.88$, with 98% credible interval [6.097, 7.691]. Compared with the basic reproduction number of SARS in 2003, that is 2.7 [8], the COVID-19 is more likely to develop in a cluster with a higher severity. Finally, we estimate the rate β_2 of susceptible individuals contacted but uninfected by COVID-19. Because the exact count of recovered

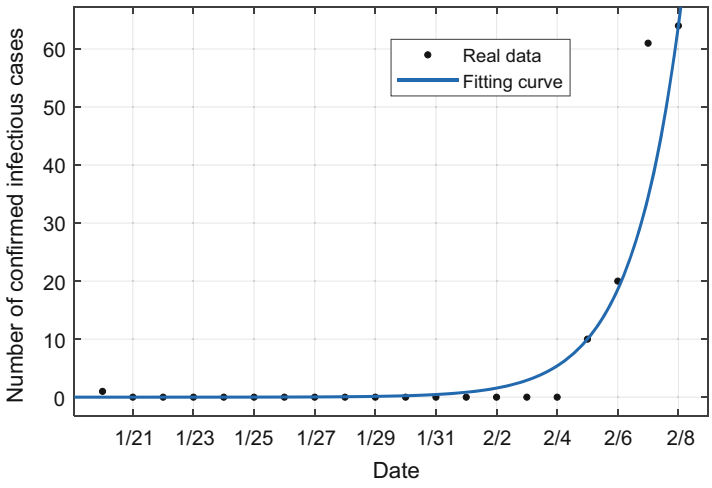


Fig. 7 Curve fitting for real data in Stage 1

individuals in the cruise ship was not observable, we apply the real data from the Health Commission of Wenzhou of China to estimate the rate β_2. The idea is based on the facts that the infection capacity of the novel coronavirus is similar anywhere, and all the required data can be collected from the Health Commission of Wenzhou [16], which is the most affected city by COVID-19 except the cities of Hubei, China. Thus, the parameter β_2 is determined by

$$\beta_2 = \frac{14096}{6475000 \times 31} = 0.00007,$$

where 14096 is the number of removed cases from medical observation during a month from January 23 to February 23, 6475000 is the urban population of Wenzhou City [17], and 31 is the number of days in a month.

We now substitute all the estimated case-specific parameters as above to the SEIR-type model governed by ODEs system (1) and use the Runge-Kutta method [2] to solve the solution of this system. Comparing to the number of infectious cases in our data set, we see that the simulated trend of the infectious cases matches well with the real trend, see Fig. 8.

Validated by the real life data, we adopt the proposed model (1) to predict the future trend of COVID-19 epidemic without quarantine intervention. Using the estimated parameter sets, we could simulate and calculate the populations of the susceptible, latent, infectious, and recovered groups at any point of time t. In the remainder of this section, we conduct three numerical simulations to answer the third and fourth questions in Sect. 2.1. Firstly, we predict the spreading trend of the novel coronavirus if no quarantine intervention is taken. As shown in Fig. 9a, we obtain the populations of all groups S, E, I, and R in Stage 1,

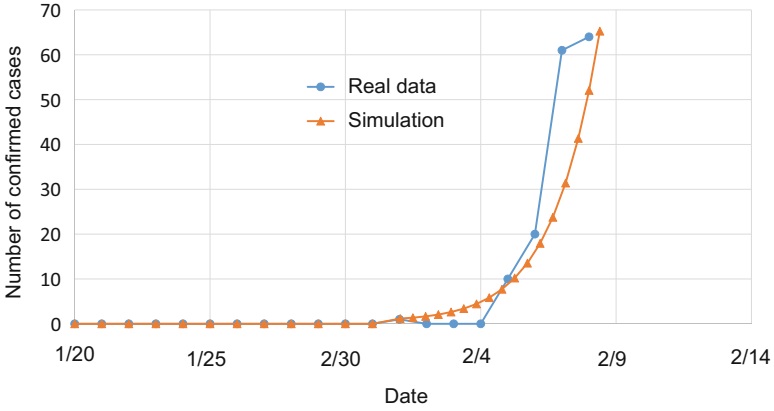


Fig. 8 Comparison between simulation and real data

among which we observe that without quarantine intervention, the coronavirus will keep spreading until the whole population were affected by COVID-19. In other words, the population of recovered group R eventually reaches 3705, and then the final scale of COVID-19, $\hat{R}$, is 0.998, which is the non-trivial solution of implicit equation (3) with the estimated parameters $\beta_1 = 0.7241$ and $\gamma = 1/9.5$. The numerical simulation result in Fig. 9a is consistent with the theoretic result of Theorem 1.

Secondly, we estimate the population of latent/exposed group at any point t. It is well-known that the latent cases have been infected by COVID-19 and actively infectious but not yet symptomatic in their incubation periods. Hence, the latent ones are quite dangerous for the entire population of a cluster community. For government officials and other decision makers, it is necessary to make clear the number of the latent cases. From Fig. 9b, we show that the population of latent group reaches its peak 15 days after the real start date of the COVID-19 outbreak in "Diamond Princess" cruise ship, and the maximum number is approximately 931. Furthermore, on the date 16 days after the COVID-19 outbreak, the infectious cases including the confirmed infectious and latent ones can get its maximum value as 1568, when the number of susceptible individuals is about 507. This yields that when $S(t) = 0.14$, the maximum scale of infectious cases $I_m(t) = 0.42$, which generally coincides with the theoretic result of Theorem 2. These simulation results shown in Fig. 9b demonstrate that the novel coronavirus can propagate quickly and infect more cases in a cluster community.

Thirdly, we calculate the total number of people that are contacted by the infectious cases at any time t. As decision makers could not observe the accurate population impacted by the COVID-19, our model proposed in this study could be used to estimate the current scope of this population, see Fig. 9c. In our model, by the time when the quarantine intervention begins to work, it is approximately

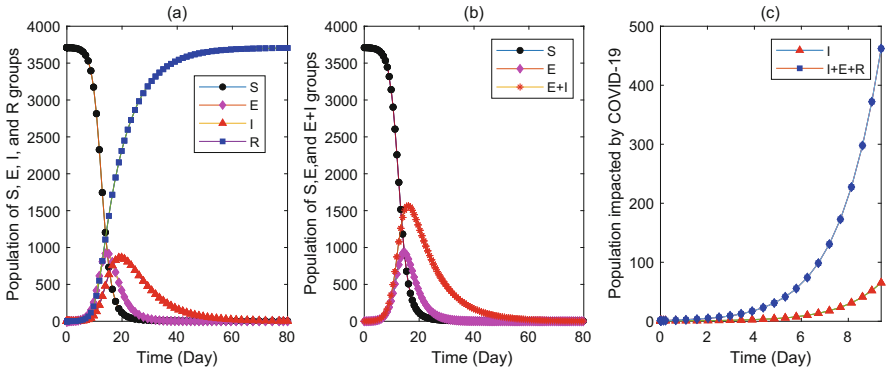


Fig. 9 Dynamics of COVID-19 in Stage 1

estimated that 462 people have been contacted by the novel coronavirus, although only 65 cases are infected.

4 Case Study and Sensitivity Analysis for Quarantine Stage

4.1 A Real World Study for Stage 2

In the second stage of the epidemic of COVID-19, the quarantine intervention is implemented to the suspectable group to prevent and control the rapid transmission of the novel coronavirus in a cluster community. On February 5, 2020, the officials of the Cruise ship "Diamond Princess" announced that the quarantine intervention was performed for 14 days. Considering that the report day is approximately 3 days after the onset date, Fig. 10 shows how the number of infectious cases vary over time from February 9 to February 21, that is the date later 2 days than February 19 on which quarantine intervention ended and passengers officially disembarked.

Similarly to Sect. 3.2, we firstly use the work-flow designed in Fig. 6 to determine the optimal case-specific parameters q, λ, σ_Q of the TQ model proposed in Stage 2. Secondly, we implement the determined set of parameters to solve the ODEs system (2). Finally, we compare the simulated solution with the real data from "Diamond Princess" to validate the proposed TQ model.

As shown in Fig. 11, we apply the curve fitting of the real data to minimize $|I(t) - I^{\text{real}}(t)|$ of Stage 2, where the dot is the data point of number of infectious cases in each time point, and the line is the fitted function curve. This curve fitting shows a good match with the real data with a goodness of fit R-square $= 0.9707$. Arguing as Sect. 3.2, it follows from Fig. 11 that the number of infectious cases in Stage 2 is approximately $\exp\{[(1-q)\beta_1 - \gamma](t+t_0)\} \approx \exp\{0.1729(t+21)\}$. Hence, the quarantine proportion is estimated as $q = 0.62$ with β_1 and γ the same as those

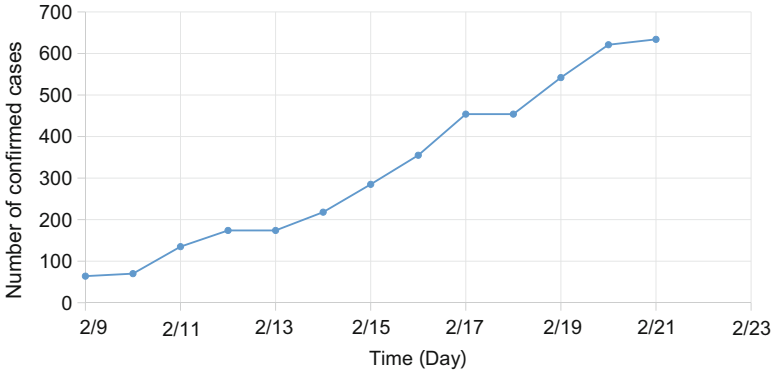


Fig. 10 Timeline of confirmed infectious cases in Stage 2

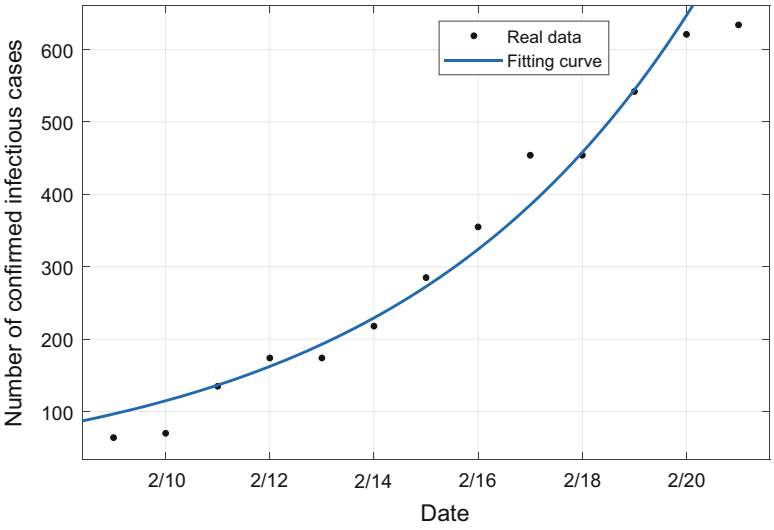


Fig. 11 Curve fitting for real data in Stage 2

in Stage 1, and $t_0 = 21$ means that the date earlier 21 days than February 9 is the real starting date of the COVID-19 outbreak, that is January 20, when the first infected passenger boarded the "Diamond Princess" cruise ship. This simulation coincides to the real data collected from this cruise ship. Following Lauer et al. [7], the median incubation period of COVID-19 was 5.1 days, then the transition rate from groups E_Q to I is $\sigma_Q = 1/5.1$. According to the requirements of health sector, the quarantine period is 14 days, then the transition rate from groups S_Q to S is $\lambda = 1/14$.

Remark 3 The further explanations for the quarantine proportion $q = 0.62$ should be emphasized from the following two aspects. On the one hand, the quarantine

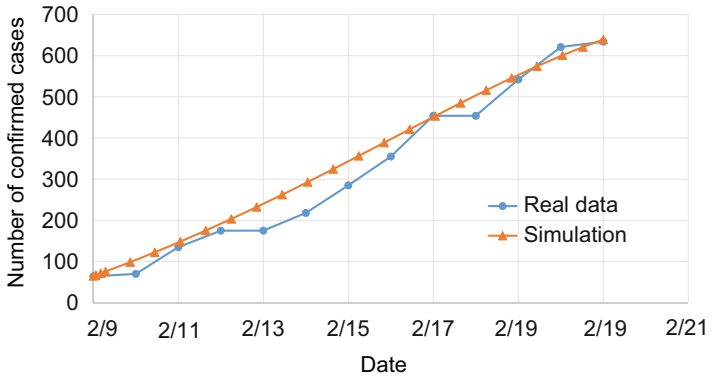


Fig. 12 Comparison between simulation and real data in Stage 2

proportion is the coverage proportion of quarantine intervention in the whole population of the susceptible group. This proportion 0.62 means that only 0.62 fraction of susceptible individuals were quarantined, the rest 0.38 fraction were still exposed to the novel coronavirus, which is due to the high cost of human and material resources for quarantine intervention. On the other hand, the quarantine proportion represents the quarantine efficacy, where $q = 0.62$ signifies that the quarantine intervention is only effective for 0.62 fraction of all quarantined susceptible individuals, namely that only 0.62 fraction of quarantined susceptible individuals move to the group S_Q, although all the passengers are quarantined on the Cruise ship "Diamond Princess." The partial efficiency of the quarantine intervention may be the poor implementation of quarantine protection measure, such as the ambiguous quarantine between fever area and non-fever area, absence of infectious disease experts, under-supply of masks and talking among passengers without masks, and walking around of fever patients. A recent New York Times highlighted breaches in quarantine measures, such as sick crew members sleeping in cabins with other members who were allowed to continue normal duties on the ship [14].

Denote by t_0 the real start time of the quarantine intervention working on the susceptible group. From the results in Stage 1, the populations of existing S, E, I, and R groups in the "Diamond Princess" at time t_0 are 3248, 223, 65, and 175, respectively. As the quarantine proportion is estimated as $q = 0.62$, the number of quarantined susceptible cases at time t_0 is 2014, and the number of remainder susceptible cases is 1234. The number of quarantined latent cases at time t_0 is 0.

We now substitute all the case-specific parameters estimated as above and those in Sect. 3.2 into the TQ model governed by ODEs system (2), and apply the Runge-Kutta method to derive the solution of this system. The comparison of the simulated curve with the estimated parameters matches well with the real world data, see Fig. 12.

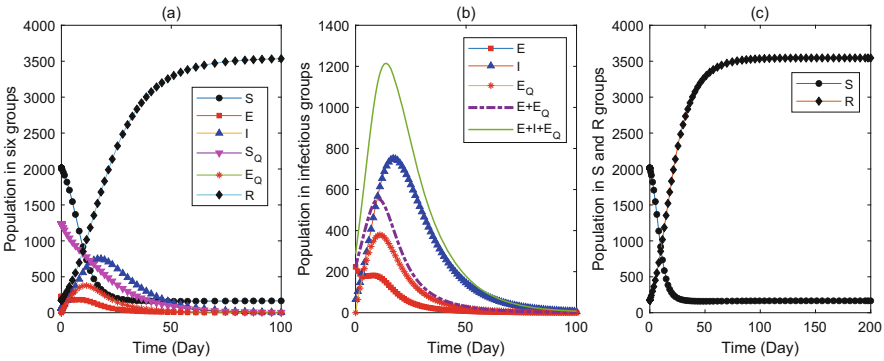


Fig. 13 Dynamics of COVID-19 in Stage 2

The TQ model in Stage 2 will help us to predict the varying trend of $S(t)$, $E(t)$, $I(t)$, $S_Q(t)$, $E_Q(t)$, and $R(t)$ under quarantine intervention. In what follows, we carry out the following two simulations of the transmission of COVID-19 in the "Diamond Princess" cruise ship. Firstly, we predict the transmission trend of the novel coronavirus if quarantine intervention is taken. As shown in Fig. 13a, we simulate and calculate the populations in all groups in Stage 2 at any point of time. In particular, Fig. 13b gives the simulated numbers of the infectious cases including the confirmed infectious, latent, and quarantined latent cases, which illustrates that with the quarantine control measure, the number of infectious cases reaches its peak value 1215 at approximately 14 days after the quarantine measure began, namely at February 23; and it takes about 10 days for the sum of numbers of latent and quarantined latent cases to reaches its maximum value 554, namely that the number of latent cases reaches its peak at February 19. These results, along with Fig. 9a, b, demonstrate that if the quarantine intervention is implemented, the fewer cases are infected at a smaller speed. In addition, we observe from Fig. 13c that as time goes by, the recovered cases will ultimately reach a population 3545, and the suspectable cases will be persistent with a population of nearly 166, which indicates that with the quarantine effort, the coronavirus does not pervade the whole population in the cruise ship.

Secondly, we conduct some numerical simulations to illustrate the effectiveness of the quarantine control strategy. See Fig. 14, we present the trend of infected cases with quarantine and without quarantine in 7 days after the quarantine intervention starts to work. Specifically, Fig. 14a demonstrates that the number of confirmed infectious cases with quarantine is 454 less than that without quarantine. Figure 14b depicts that the number of latent cases at the 7-th day with quarantine is 475 less than that without quarantine. From Fig. 14c, we see that the number of the infectious cases (including those in I, E, and E_Q groups) with quarantine is 928 less than that without quarantine. The conclusions above indicate that with quarantine control, the number of confirmed infectious, latent, and infectious cases in 7 days will significantly decrease by about 60%, 58%, and 59%, respectively, compared

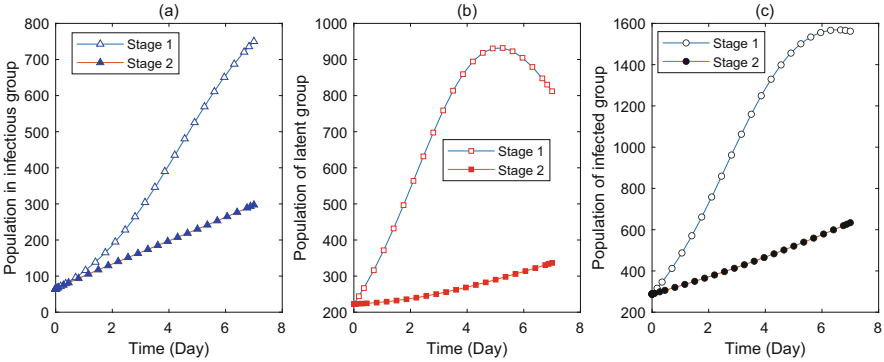


Fig. 14 Dynamics with quarantine and without quarantine

with the scenario of no quarantine control. Hence, we conclude that introducing a quarantine control into a cluster community is an effective tool to defend against the outbreak of COVID-19.

4.2 Sensitivity Analysis

Sensitivity analysis is applied to identify the criticality of the modeling parameters and the effectiveness of quarantine intervention during the COVID-19 outbreak. So in the subsection, we use sensitivity analysis as a tool to explore how different parameters and quarantine strategies will impact the disease control effectiveness. Specifically, we aim to study how the final scope and the duration of COVID-19 are impacted by various factors, including but not limited to the COVID-19 popularity, and the number of initial quarantined/infectious cases.

Firstly, we compare the estimated final scope of the epidemic of COVID-19 with quarantine and without quarantine, and how this final scope varies over the disease infection rate β_1. As shown in Fig. 15, the final scope of the COVID-19 epidemic is hardly variable between the conditions with quarantine and without quarantine when the novel coronavirus is not very popular, namely when $\beta_1 < 0.1067$. As the popularity of the virus increases, namely $\beta_1 > 0.1067$, the difference of the numbers of recovered cases between the conditions with and without quarantine firstly increases exponentially, which means the quarantine intervention plays a crucial role to prevent and control the serious COVID-19 outbreak.

Secondly, we investigate how the end time of the outbreak of COVID-19 varies over the number of initial infectious cases, where the end time is defined as the time point when the number of infectious cases is less than one. From Fig. 16, the duration of the COVID-19 epidemic increases concavely as the number of initial infectious cases increases.

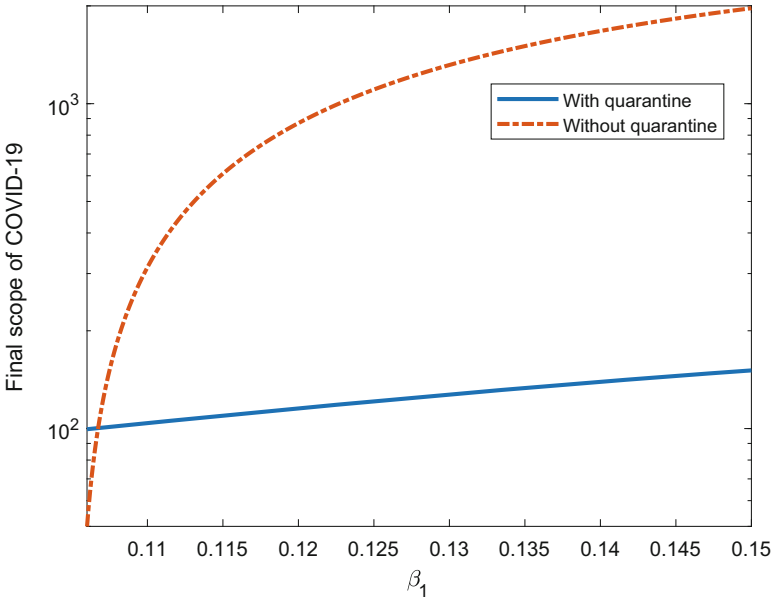


Fig. 15 Final scope of COVID-19 with and without quarantine

Thirdly, we compare how the number of infectious cases and the final scope of COVID-19 epidemic vary over the different numbers of quarantined suspectable individuals at the start time of quarantine intervention. As shown in Fig. 17a, as the number of quarantined suspectable individuals increases, the number of infectious cases decreases much more rapidly. From Fig. 17b, it follows that the final scope of COVID-19 outbreak decreases with the increase of the number of quarantined suspectable individuals. Hence by Fig. 17a, b, we conclude that the more initial number of quarantined suspectable individuals, the less the impact of the novel coronavirus on the susceptible individuals and the earlier the COVID-19 outbreak ends. As government officials or other decision makers would like to get rid of the influence of the COVID-2019 as soon as possible, we now turn to study the end time of the COVID-19 epidemic. The simulation of Fig. 17c illustrates that the end time is decreasing with increasing marginal returns while adding more quarantined cases, which is consistent with the results in Fig. 17a, b.

The massive quarantine intervention comes with large economic costs and social disruption effects, which prompts a question of what optimal quarantine strategy can be identified to restart economic and social activities as well as minimize the burden of the epidemic. To this end, we study what fraction of the population would be quarantined under the proposed contact-tracing and quarantine strategy. As shown in Fig. 17c, we observe that from the quarantine number 1541 on, the end time of the COVID-19 epidemic decrease sharply and have no recurrence. Hence, the optimal number of quarantined individuals is 1541, about 41.5% of the population

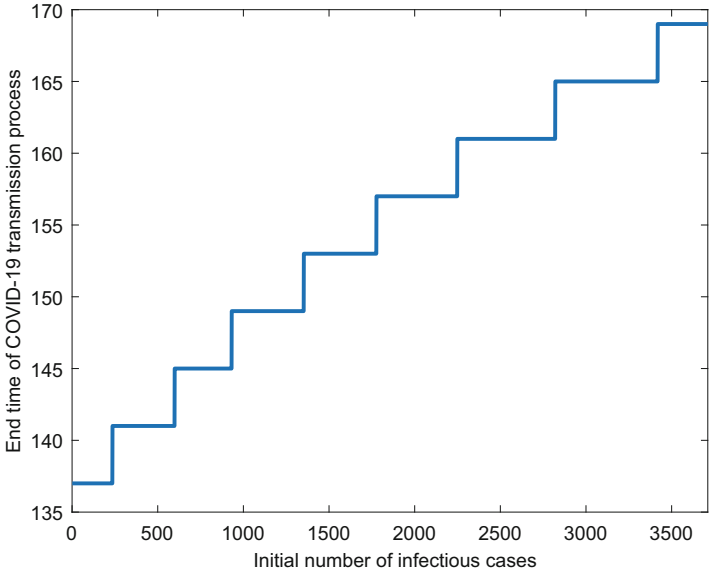


Fig. 16 Duration of COVID-19 outbreak varies over initial infectious numbers

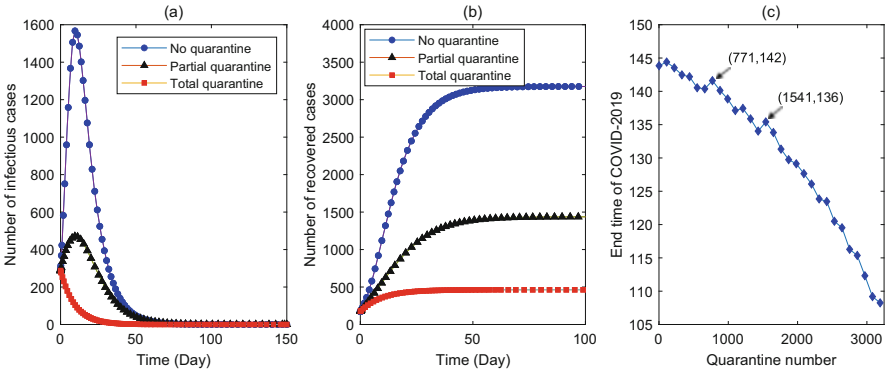


Fig. 17 Impact of initial number of quarantined individuals

3711 in the "Diamond Princess" cruise ship, and similarly the sub-optimal number is 771, about 21%, which are certainly relevant fractions of the population, and much better options compared with massive quarantine intervention affecting the whole population in a long run. As a result, governments and decision makers could determine the intensity of quarantine intervention according to the degree of close contact with the infectious cases, so as to balance the COVID-19 epidemic control and the economic and social development. Precisely, the fraction 41.5% (or 21%) of the entire population in a cluster community is the optimal (or sub-optimal)

quarantine proportion implemented to the exposed contacts according to the degree of contact from more close to less close.

5 Discussion

Coronavirus Disease 2019 (COVID-19) is a recently described zoonotic illness which has spread rapidly and widely throughout the world since December 2019 and is identified as a global pandemic by the World Health Organization. The causative agent of COVID-19 is a novel coronavirus, which is mainly transmitted by close aerosol and close contact and has a significant aggregation phenomenon in cluster communities. The outbreak of COVID-19 has attracted considerable global attention. Public-health interventions to prevent and control the transmission are expected to have a substantial impact on reducing the size of the epidemic, among which an effective quarantine measure is a potential tool to achieve the maximal protection of population with the minimal interruption of social-economic activities.

In this chapter, we use an epidemiological method by ordinary differential equations system to study the transmission dynamics and quarantine intervention of the COVID-19 outbreak in a cluster community. We divide the whole life-cycle of the COVID-19 epidemic into two stages. The first stage is emerging stage, which refers to the time period between the coronavirus start point and the official quarantine point. The second stage refers to the time period between the quarantine point and the epidemic end point, namely that the point when all the infected population are recovered, where the quarantine intervention implemented can reduce the risk of exposure of individuals to the novel coronavirus, and then the magnitude and the speed of COVID-19 outbreak can decrease.

In the first stage before quarantine effort, we consider the transmission dynamics of COVID-19 in a cluster community. Before the appearance of the quarantine intervention, we divide the whole population into susceptible (S), exposed (E), infectious (I), and recovered (R) groups and then propose a general SEIR-type epidemic model. We analytically study the epidemiological parameters of the COVID-19 transmission, including the final scope of COVID-19 epidemic, the maximal scale of infectious cases, and some threshold values which play a critical role to control the epidemic of COVID-19. In the real world case study, we use the epidemiologic data from the "Diamond Princess" cruise ship, design a work-flow to determine the optimal case-specific parameters, validate the proposed model by comparing the simulated curve with the real data, and estimate by our model the varying trend and the popularity of COVID-19 epidemic without quarantine measure.

In the second stage with quarantine intervention, we construct a novel Transmission-Quarantine model to explore the interplay mechanism between the transmission dynamics and quarantine intervention of the COVID-19 epidemic. We solve the system and compare them with the real data from the Cruise ship "Diamond Princess." The real data matches well with our predicted solution. We

predict the transmission trend of COVID-19 under quarantine intervention, which indicates that if the quarantine intervention is taken, the fewer cases are infected at a smaller speed, and the new coronavirus does not pervade the whole population in a cluster community. Also, we present the trend of infected cases with quarantine and without quarantine in 7 days after the quarantine intervention starts to work, which illustrates the effectiveness of the quarantine control strategy, namely that with quarantine intervention, the number of infectious cases in 7 days decreases by about 60%, compared with the scenario of no intervention.

Moreover, we apply sensitivity analysis to identify the criticality of the modeling parameters and the effectiveness of quarantine intervention during the COVID-19 outbreak. Firstly, the quarantine intervention is not necessary when the novel coronavirus is not very popular with $\beta_1 < 0.1067$. However, when $\beta_1 > 0.1067$, the quarantine intervention is critical and could make a significant influence on the final scope of COVID-19 epidemic. Secondly, the duration of COVID-19 epidemic increases concavely as the initial number of infectious cases increases. Thirdly, the more initial number of quarantined suspectable cases, the less the impact of the novel coronavirus on the susceptible individuals and the earlier the COVID-19 outbreak ends; and the end time decreases with diminishing marginal returns while adding more quarantined cases. However, the quarantine intervention is at the high price of economic loss and social disruption. Governments and policy makers should determine the intensity of quarantine intervention according to the degree of close contact with the infectious cases, so as to balance the COVID-19 epidemic and the economic and social activities. Precisely, the fraction 41.5% (or 21%) of the entire population in a cluster community is the optimal (or sub-optimal) quarantine proportion implemented to the exposed contacts according to the degree of contact from more close to less close.

Acknowledgment The research was supported by the National Social Science Foundation of China (No.21BTJ029).

Appendix: Proofs of Theorems

Proof of Theorem 1 Consider the non-linear ODEs system (1), we divide the last equation by the first one to have that

$$\frac{dR(t)}{dS(t)} = -\frac{\beta_2}{\beta_1+\beta_2} - \frac{\gamma}{(\beta_1+\beta_2)S(t)},$$

namely

$$dR(t) = -\left(\frac{\gamma}{(\beta_1+\beta_2)S(t)} + \frac{\beta_2}{\beta_1+\beta_2}\right) dS(t).$$

Then, we take the integral on both sides of the above equation to obtain that

$$R(t) = -\frac{\gamma}{\beta_1 + \beta_2} \ln S(t) - \frac{\beta_2}{\beta_1 + \beta_2} S(t) + C, \tag{4}$$

where C is a constant determined by the initial conditions:

$$S(0) = \frac{N-1}{N} \approx 1, \quad I(0) = \frac{1}{N} \approx 0, \quad E(0) = 0 \quad \text{and} \quad R(0) = 0,$$

following from the fact that there is only one infectious case who starts the transmission of COVID-19 in the "Diamond Princess" cruise ship. Now we substitute these initial conditions into (4) to get

$$C = \frac{\beta_2}{\beta_1 + \beta_2}.$$

This, along with (4), implies that $R(t)$ has the explicit expression with respect to $S(t)$ as

$$R(t) = -\frac{\gamma}{\beta_1 + \beta_2} \ln S(t) - \frac{\beta_2}{\beta_1 + \beta_2} S(t) + \frac{\beta_2}{\beta_1 + \beta_2}, \tag{5}$$

As the end point of the COVID-19 outbreak is defined as the time point when the number of the infectious cases is less than one, we observe that $E(\infty) + I(\infty) = 0$, which, along with $S(t) + E(t) + I(t) + R(t) = 1$ for time $t \geq 0$, yields that $S(\infty) = 1 - R(\infty) = 1 - \hat{R}$. Thus from (5), it follows that when $t \to \infty$,

$$\hat{R} = -\frac{\gamma}{\beta_1 + \beta_2} \ln(1 - \hat{R}) - \frac{\beta_2}{\beta_1 + \beta_2}(1 - \hat{R}) + \frac{\beta_2}{\beta_1 + \beta_2}.$$

After a simple computation, we have

$$\ln(1 - \hat{R}) = -\frac{\beta_1}{\gamma} \hat{R},$$

which prove that the final scope $\hat{R}$ of COVID-19 satisfies the implicit equation (3) in Theorem 1.

Obviously, $\hat{R} = 0$ is a trivial solution of Eq. (3). Now it suffices to prove the existence of the non-trivial solution $\hat{R}$, $0 < \hat{R} < 1$, under the condition that $\frac{\beta_1}{\gamma} > 1$. Setting $f(x) = 1 - x - e^{-\frac{\beta_1}{\gamma} x}$, $0 < x < 1$, and taking the derivative of $f(x)$ with respect to x, we have that $f'(x) = \frac{\beta_1}{\gamma} e^{-\frac{\beta_1}{\gamma} x} - 1$ and $f''(x) = -\left(\frac{\beta_1}{\gamma}\right)^2 e^{-\frac{\beta_1}{\gamma} x} < 0$, which yields that $f(x)$ is a concave function, see Fig. 18. Letting $f'(x_0) = 0$, we have $x_0 = \frac{\gamma}{\beta_1} \ln \frac{\beta_1}{\gamma}$, where $x_0 \in (0, 1)$ since $\frac{\beta_1}{\gamma} > 1$. Moreover, $f(1) = -e^{-\frac{\beta_1}{\gamma}} < 0$,

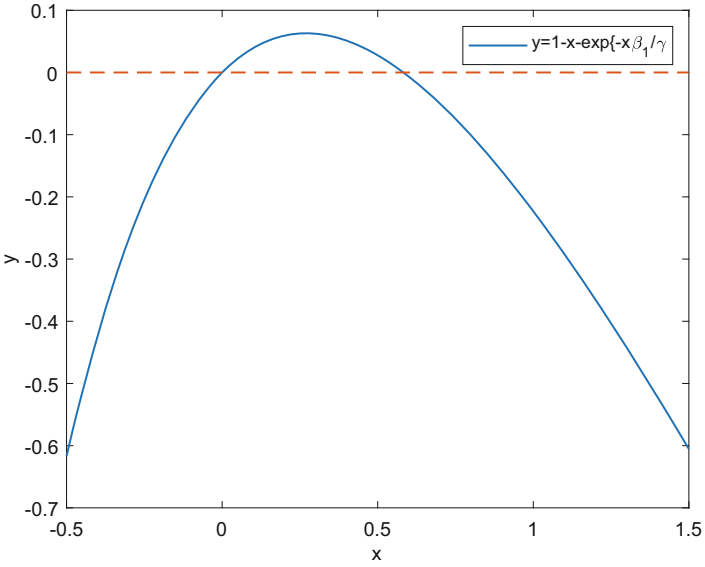


Fig. 18 Curve of $y = 1 - x - \exp\{x\beta_1/\gamma\}$

and $f(x_0) = \frac{\gamma}{\beta_1}\left(\frac{\beta_1}{\gamma} - 1 - \ln\frac{\beta_1}{\gamma}\right)$, where $f(x_0) > 0$ since $\ln x < x - 1$ for $x \in (0, 1)$. By the intermediate value theorem, the equation $f(x) = 0$ has a non-trivial solution $\hat{R}$ such that $0 < x_0 < \hat{R} < 1$, which shows that when $\frac{\beta_1}{\gamma} > 1$, there exists the claimed non-trivial solution $\hat{R}$, $0 < \hat{R} < 1$. □

Taking into account the numerical values used in Sect. 3.2, namely that $\beta_1 = 0.8643$ and $\gamma = 0.1$, we conduct the numerical simulation to solve the Eq. (3), see Fig. 19. This illustrates that $\hat{R}$ has two solutions: zero and non-trivial solution, of which the numerical solution is nearly 1. The simulation result is consistent with the real data from the Cruise ship "Diamond Princess."

Proof of Theorem 2 Without loss of generality, we only prove that $\hat{R}$ increases as β_1 increases. Taking the partial derivative of $\hat{R}$ with respect β_1 on the both sides of Eq. (3), we have

$$\frac{\partial \hat{R}}{\partial \beta_1} = \exp\left\{-\frac{\beta_1}{\gamma}\hat{R}\right\}\left(\frac{\hat{R}}{\gamma} + \frac{\beta_1}{\gamma}\frac{\partial \hat{R}}{\partial \beta_1}\right),$$

which yields that

$$\frac{\partial \hat{R}}{\partial \beta_1} = \frac{\frac{\hat{R}}{\gamma}\exp\left\{-\frac{\beta_1}{\gamma}\hat{R}\right\}}{1 - \frac{\beta_1}{\gamma}\exp\left\{-\frac{\beta_1}{\gamma}\hat{R}\right\}}.$$

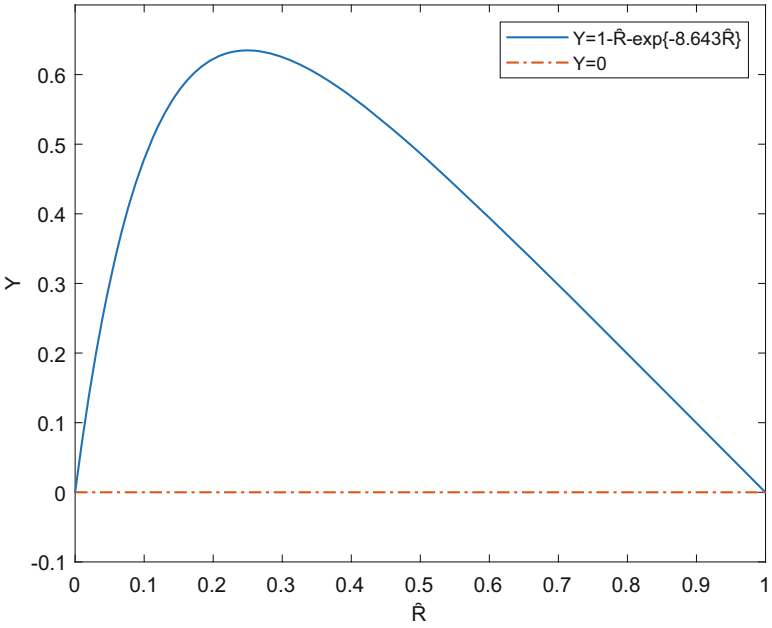


Fig. 19 Curves of Y = 1-R̂-exp{−8.643R̂} and Y = 0

Let $u = \frac{\beta_1}{\gamma}$, where $u > 1$ following from Theorem 1. So it remains to show that $1 - ue^{-u\hat{R}} > 0$. Note that $\hat{R}$ is the non-trivial solution of Eq. (3) and also is a function of β_1. Hence, the point $\hat{R}$ is the point on the horizontal axis corresponding to the intersection of line $y = x$ and curve $y = 1 - e^{-ux}$. As shown in Fig. 20, since $1 - e^{-u\hat{R}} > 1 - ue^{-u\hat{R}}$, curve $y = 1 - e^{-u\hat{R}}$ is above curve $y = 1 - ue^{-u\hat{R}}$; and curve $y = 1 - ue^{-u\hat{R}}$ cross the horizontal axis at $x = \frac{\ln u}{u}$. Thus we can show that $1 - ue^{-u\hat{R}} > 0$, if we prove that on the horizontal axis point $x = \frac{\ln u}{u}$ is on the left side of $x = \hat{R}$, in other words, $1 - e^{-u(\ln u)/u} > \frac{\ln u}{u}$, i.e., $1 - \frac{1}{u} - \frac{\ln u}{u} > 0$. In fact, we define a function $f(u) = 1 - \frac{1}{u} - \frac{\ln u}{u}$ and obtain that $f(u) > 0$ for any $u > 1$, which is due to $f'(u) = \frac{\ln u}{u^2} > 0$ and $f(1) = 0$. □

Proof of Theorem 4 Summing the second and third equations of system (1), and then diving the sum by the first equation, we have

$$\frac{d(E(t) + I(t))}{dS(t)} = \frac{\gamma}{(\beta_1 + \beta_2)S(t)} - \frac{\beta_1}{\beta_1 + \beta_2},$$

that is,

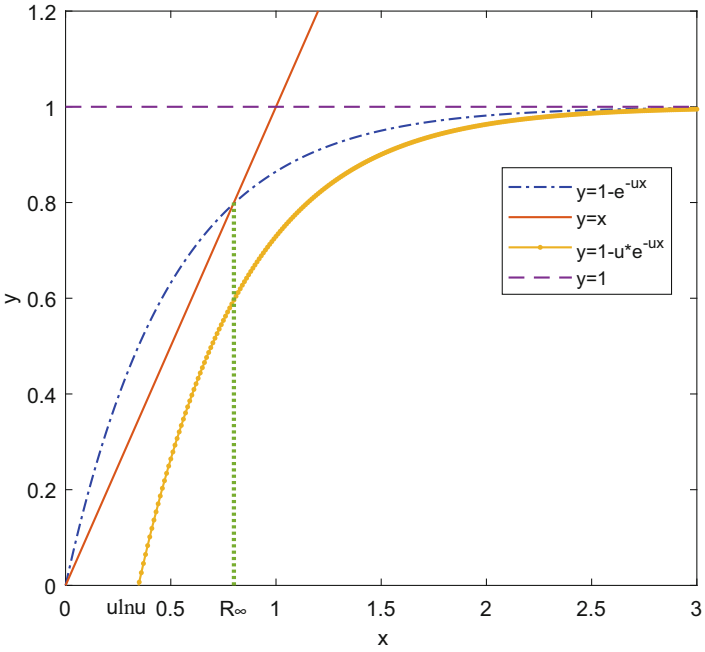


Fig. 20 Illustration of the proof of Theorem 2

$$d(E(t)+I(t)) = \left(\frac{\gamma}{(\beta_1+\beta_2)S(t)} - \frac{\beta_1}{\beta_1+\beta_2}\right) dS(t).$$

Now we take the integral on both sides of the above equation to obtain that

$$E(t)+I(t) = -\frac{\beta_1}{\beta_1+\beta_2}S(t) + \frac{\gamma}{\beta_1+\beta_2}\ln S(t) + \bar{C},$$

where the constant $\bar{C}$ is determined by the initial conditions as those in the proof of Theorem 1. So we get $\bar{C} = \beta_1/(\beta_1+\beta_2)$. Further, we derive that $E(t)+I(t)$ satisfy the explicit expression with respect to $S(t)$ as

$$E(t)+I(t) = \frac{\gamma}{\beta_1+\beta_2}\ln S(t) - \frac{\beta_1}{\beta_1+\beta_2}S(t) + \frac{\beta_1}{\beta_1+\beta_2}. \tag{6}$$

From (6), letting $\frac{d(E(t)+I(t))}{dS(t)} = 0$ implies that

$$S(t) = \frac{\gamma}{\beta_1},$$

at which we get the maximum value of $E(t)+I(t)$ as:

$$I_{\max}(t) = \frac{\gamma}{\beta_1 + \beta_2}\left(\ln\frac{\gamma}{\beta_1} - 1\right) + \frac{\beta_1}{\beta_1 + \beta_2},$$

which is the maximum scale of the infectious cases (including the latent ones and the confirmed ones). □

Considering the numerical values $\beta_1 = 0.8643$, $\beta_2 = 0.00007$, and $\gamma = 0.1$, used in Sect. 3.2, we implement the numerical simulation for the Eq. (6), shown in Fig. 21. It can be seen that the fraction of infectious cases reaches its curve peak 0.6347 when the percentage of susceptible individuals is about 0.1141, which coincides with the real data from the "Diamond Princess" Cruise ship.

Proof of Theorem 5 Using (5) and letting

$$\frac{dR(t)}{dS(t)} = -\frac{\beta_2 S(t) + \gamma}{(\beta_1 + \beta_2)S(t)} \geq -1,$$

we have

$$S(t) \geq \frac{\gamma}{\beta_1},$$

whereas, letting $\frac{dR(t)}{dS(t)} < -1$ leads to

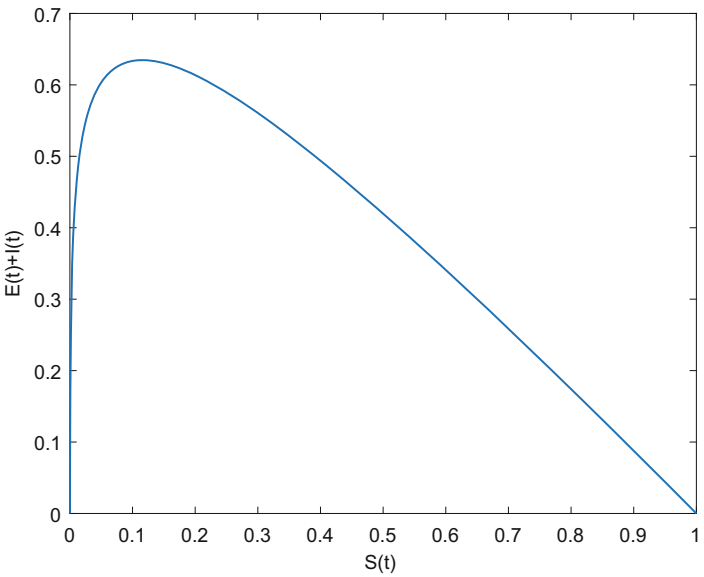


Fig. 21 Curve of E(t)+I(t) with respect to S(t)

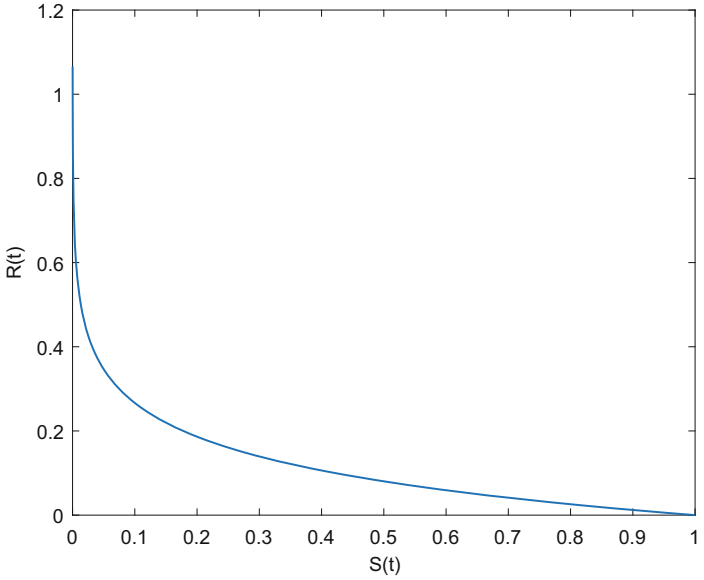


Fig. 22 Curve of R(t) with respect to S(t)

$$S(t) < \frac{\gamma}{\beta_1}.$$

Then we conclude that when $S(t) \geq \frac{\gamma}{\beta_1}$, the scale of recovered cases, $R(t)$, increases slowly as $S(t)$ decreases; while when $S(t) < \frac{\gamma}{\beta_1}$, $R(t)$ increases sharply as $S(t)$ decreases. □

The function $R(t)$ with respect to $S(t)$ is expressed as (5), and its curve of $R(t)$ with respect to $S(t)$ is shown in Fig. 22, where we see that the curve of $R(t)$ is relatively gentle when $S(t) \geq 0.1157$, whereas the curve of $R(t)$ becomes considerable steep when $S(t) < 0.1157$.

References

1. N. Bellomo, R. Bingham, M.A.J. Chaplain, G. Dosi, G. Forni, D.A. Knopoff, J. Lowengrub, R. Twarock, M.E. Virgillito, A multi-scale model of virus pandemic: heterogeneous interactive entities in a globally connected world. Math. Models Methods Appl. S. **30**(8), 1591–1651 (2020)
2. J.C. Butcher, *The Numerical Analysis of Ordinary Differential Equations: Runge-Kutta and General Linear Methods* (Wiley, New York, 1987)
3. Y. Chen, Q. Liu, D. Guo, Coronaviruses: genome structure, replication, and pathogenesis. J. Med. Virol. **92**(4), 418–423 (2020)

4. Coronavirus disease (COVID-19) outbreak situation. https://www.who.int/emergencies/diseases/novel-coronavirus-2019
5. COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). https://coronavirus.jhu.edu/map.html
6. W.O. Kermack, A.G. Mckendrick, Contribution to the mathematical theory of epidemics. Proc. R. Soc. Lond. A Containing Pap. Math. Phys. Character **115**(772), 700–721 (1927)
7. S.A. Lauer, K.H. Grantz, Q. Bi, F.K. Jones, Q. Zheng, H.R. Meredith, A.S. Azman, N.G. Reich, J. Lessler, The incubation period of Coronavirus Disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. Ann. Intern. Med. **172**(9), 577–582 (2020)
8. M. Lipsitch, T. Cohen, B. Cooper et al., Transmission dynamics and control of severe acute respiratory syndrome. Science **300**(5627), 1966–1970 (2003)
9. Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong et al., Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N. Engl. J. Med. **382**, 1199–1207 (2020)
10. Medical treatment and scientific research of severe patients. Press conference of National Health Commission of China on 4 Feb 2020 (in Chinese). http://www.nhc.gov.cn/xwzb/webcontroller.do?titleSeq=11220&gecstype=1
11. Ministry of Health, Labour and Welfare of Japan. Number of COVID-19 cases with symptoms among the passengers on the Cruise ship (by close contact status, as of 19 Feb 2020). https://www.mhlw.go.jp/content/10200000/Fig2.pdf
12. S. Riley, C. Fraser, C. Donnelly, A. Ghani et al., Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions. Science **300**(5627), 1961–1966 (2003)
13. Situation Updates-SARS: Update 95-Chronology of a serial killer 2003. https://www.who.int/csr/don/2003_06_18/en/
14. The New York Times, We're in a Petri Dish': How a Coronavirus Ravaged a Cruise Ship, 22 Feb 2020. https://www.nytimes.com/2020/02/22/world/asia/coronavirus-japan-cruise-ship.html?searchResultPosition=1
15. The Official Website of Diamond Princess: Cruises P, Princess Cruises: Diamond Princess Coronavirus & Quarantine Updates-Notices & Advisories Princess Cruises website:@ Princess Cruises; 2020. https://www.princess.com/news/notices_and_advisories/notices/diamondprincess-update.html
16. The Official Website of the Health Commission of Wenzhou (in Chinese). http://wjw.wenzhou.gov.cn/
17. The Population of Wenzhou City (in Chinese). http://www.wenzhou.gov.cn/col/col1217817/index.html
18. The State Council Information Office of the People's Republic of China, Fighting COVID-19: China in Action, June 2020. http://www.scio.gov.cn/zfbps/32832/Document/1681809/1681809.htm
19. H. Tian, Y. Liu, Y. Li, C.H. Wu, B. Chen et al., An investigation of transmission control measures during the first 50 days of the COVID-19. Science **368**(6491), 638–642 (2020)
20. World Health Organization (WHO). Coronavirus. https://www.who.int/health-topics/coronavirus. Accessed 23 Jan 2020
21. Z. Yang, Z. Zeng, K. Wang, S.S. Wong, W. Liang, M. Zanin, P. Liu et al., Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. J. Thorac. Dis. **12**(3), 165–174 (2020)
22. P. Zhou, X.L. Yang, X.G. Wang, B. Hu, L. Zhang, W. Zhang et al., A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature **579**, 270–273 (2020)

A 2D Kinetic Model for Crowd Dynamics with Disease Contagion

Daewa Kim and Annalisa Quaini

1 Introduction

The goal of this chapter is to extend to 2D a 1D model presented in [29] to study the early stage of an infectious disease spreading in an intermediate size population occupying a confined environment, e.g., a ER wait room, for a short period of time (minutes or hours). In this context, classical epidemiological models fail as they rely on averaged large population behaviors over a long time span (weeks or months). We focus on a kinetic theory approach [9, 12] for a disease that it spreads with close proximity of individuals, e.g., measles, influenza, or COVID-19. In the first part of the chapter, we present a model for the disease spreading based on the simplifying assumption that people's walking speed and direction are given. In the second part, ideas from the simplified model are used to incorporate disease spreading in a kinetic theory approach for crowd dynamics.

The COVID-19 pandemic has motivated a large body of literature in mathematical epidemiology. Several possible models have been investigated, including individual-based Markov models (see, e.g., [20]), Susceptible, Exposed, Infectious, or Recovered (SEIR) models (see, e.g., [25]), and networks of nodes where each mode dynamics has a SEIR structure (see, e.g., [15, 21]). Some works (e.g., [31]) model the evolution of the COVID-19 epidemics in combination with other phenomena, such as fear. Other works focus a multiscale modeling approach. A notable example that features an interdisciplinary framework including applied

D. Kim
Department of Mathematics, West Virginia University, Morgantown, WV, USA
e-mail: daewa.kim@mail.wvu.edu

A. Quaini (✉)
Department of Mathematics, University of Houston, Houston, TX, USA
e-mail: quaini@math.uh.edu

N. Bellomo, M. A. J. Chaplain (eds.), *Predicting Pandemics in a Globally Connected World, Volume 1*, Modeling and Simulation in Science, Engineering and Technology, https://doi.org/10.1007/978-3-030-96562-4_9

mathematics, immunology, economics, and virology is [11]. Similarly, modeling and simulation of disease spreading in pedestrian crowds has recently become a topic of increasing relevance. In [19], contact evolution is combined with a stochastic infection spread model to simulate disease spreading in queuing pedestrians. Agent-based numerical simulations of pedestrian dynamics are used in [23] to assess the behavior of pedestrians in public places in the context of contact transmission of infectious diseases and gather insights about exposure times and the overall effectiveness of distancing measures. Finally, we mention the combination of a microscopic force-based model with a contact tracking method in [32] to simulate the initial spreading of a highly infectious airborne disease in a confined environment.

The model that we propose for disease spreading in a walking crowd takes inspiration from the work on emotional contagion (i.e., spreading of fear or panic) in Ref. [13, 36]. In order to focus on the spreading mechanism, we first introduce a simplifying assumption: the walking speed and direction are given. The key features of the resulting simplified model are:

– A variable that denotes the level of exposure to people spreading the disease, with the underlying idea that the more a person is exposed the more likely they are to get infected
– A model parameter that describes the contagion interaction strength
– A kernel function that is a decreasing function of the distance between a person and a spreading individual

We show preliminary results for a problem involving a small crowd in a square walkable domain with no obstacles or walls.

In the second part of the chapter, the simplifying assumption is removed: the walking speed and direction result from a kinetic model of crowd dynamics, which incorporates a term to account for disease spreading. Kinetic models are derived from observing the system at the mesoscale (see, e.g., [1, 5–10]), i.e., the intermediate scale between the macroscopic one (see, e.g., [17, 26, 34]) and the microscopic one (see, e.g., [3, 16, 18, 24]). The framework for mesoscale models comes from the kinetic theory of gases, with the main difference that interactions of "active" particles are irreversible, non-conservative, and, in some cases, non-local and nonlinearly additive [2]. In general, one derives the kinetic model from interactions at the microscopic scale and the same modeling principles can lead to macroscopic (or hydrodynamic) models. See, e.g., Ref. [4]. We remark that the approach to model crowd dynamics with disease contagion presented here is different from the one in Ref. [28], where two models (one for pedestrian dynamics and the other for disease spreading) are coupled.

The particular kinetic model for crowd dynamics that we extend to account for disease spreading was first presented in [27]. Based on earlier works [1, 6], the model's main features are:

– Discrete walking directions
– Interactions modeled using tools of stochastic games

– Heuristic, deterministic modeling of the walking speed corroborated by experimental evidence [33]

In [27], the model has been shown to compare favorably against experimental data related to egression from a facility of a medium-sized group of people (40 to 138 pedestrians) [35]. In addition, we demonstrated that realistic scenarios, such as passengers moving through one terminal of Hobby Airport in Houston (USA), can be reproduced [29].

The chapter is organized as follows. We introduce our simplified contagion model in Sect. 2 and its numerical discretization in Sect. 3. Numerical results are shown in Sect. 4. Section 5 presents the kinetic theory approach for crowd dynamics extended to incorporate disease spreading. Conclusions are drawn in Sect. 6.

2 A Simplified Two-Dimensional Kinetic Model

We start from an agent-based model. At the microscopic level, we consider a group of N people divided between healthy or not spreading the disease yet (N_h) and actively spreading (N_s), with $N = N_h + N_s$. If person n belongs to the group of healthy (or not spreading) people, we denote with $q_n \in [0, 1)$ their level of exposure to people spreading the disease, with the underlying idea that the more a person is exposed the more likely they are to get infected. If person n is actively spreading the disease, then $q_n = 1$ and this value stays constant throughout the entire simulation time.

Let $[t_0, t_f]$ be a time interval of interest. Let $\boldsymbol{x}_n(t) = (x_n(t), y_n(t))^T$ and $v_n(t)$ denote the position and speed of person n for $t \in [t_0, t_f]$, with initial position $\boldsymbol{x}_n(t_0)$ and initial speed $v_n(t_0)$ being given. The microscopic model reads for $n = 1, 2, 3, \ldots, N$:

$$\frac{d\boldsymbol{x}_n}{dt} = v_n(\cos\theta, \sin\theta)^T, \quad \frac{dq_n}{dt} = \gamma \max\{(q_n^* - q_n), 0\}, \quad q_n^* = \frac{\sum_{m=1}^{N} \kappa_{n,m} q_m}{\sum_{m=1}^{N_s} \kappa_{n,m}}, \tag{1}$$

where the walking speed v_n and walking direction θ are assumed to be given. In model (1), q_n^* corresponds to a weighted average "level of sickness" surrounding person n, with $\kappa_{n,m}$ serving as the weight in the average. We define $\kappa_{n,m}$ as follows:

$$\kappa_{n,m} = \begin{cases} \kappa(|\boldsymbol{x}_n - \boldsymbol{x}_m|) = \dfrac{R}{(|\boldsymbol{x}_n - \boldsymbol{x}_m|^2 + R^2)\pi} & \text{if } m \text{ is spreading,} \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

Notice that if person m is spreading the disease, the interaction kernel is a decreasing function of mutual distance between two people and is parametrized by

an interaction distance R. The value of R is set so that the value of $\kappa_{n,m}$ is "small" at about 6 ft or 2 m. More details are given in Sect. 4. The meaning of parameter γ in (1) is contagion strength: for $\gamma = 0$ there is no contagion, while for $\gamma \neq 0$ the contagion occurs and it is faster for larger values of γ. Note that obviously the level of exposure can only increase over time. In addition, the second equation in (1) ensures that the people spreading the disease will constantly have $q_n = 1$ in time.

In order to derive a kinetic model from the agent-based model (1), we introduce the empirical distribution:

$$h^N = \frac{1}{N}\sum_{n=1}^{N} \delta(\boldsymbol{x} - \boldsymbol{x}_n(t))\delta(q - q_n(t)),$$

where δ is the Dirac delta measure. We assume that all people remain in a fixed compact domain Ω for the entire time interval under consideration: $(\boldsymbol{x}_n(t), q_n(t)) \in \Omega \subset \mathbb{R}^3$, for $n = 1, 2, 3, \ldots, N$ and $t \in [t_0, t_f]$. Let $\mathcal{P}(\mathbb{R}^3)$ be the space of probability measures on $\mathbb{R}^3$. The sequence $\{h^N\}$ is relatively compact in the weak* sense (see, e.g., [14]). Therefore, there exists a subsequence $\{h^{N_k}\}_k$ such that h^{N_k} converges to h with weak*-convergence in $\mathcal{P}(\mathbb{R}^3)$ and pointwise convergence in time as $k \to \infty$.

Let $\psi \in C_0^1(\mathbb{R}^3)$ be a test function. We have

$$\begin{aligned}
\frac{d}{dt}\langle h^N, \psi\rangle_{\boldsymbol{x},q} &= \frac{d}{dt}\left\langle \frac{1}{N}\sum_{n=1}^{N} \delta(\boldsymbol{x} - \boldsymbol{x}_n(t))\delta(q - q_n(t)), \psi \right\rangle_{\boldsymbol{x},q} \\
&= \frac{d}{dt}\frac{1}{N}\sum_{n=1}^{N} \psi(\boldsymbol{x}_n(t), q_n(t)) \\
&= \frac{1}{N}\sum_{n=1}^{N} \left(\psi_x v_n \cos\theta + \psi_y v_n \sin\theta + \psi_q \gamma \max\{(q_n^* - q_n), 0\}\right) \\
&= \langle h^N, \psi_x v \cos\theta\rangle_{\boldsymbol{x},q} + \langle h^N, \psi_y v \sin\theta\rangle_{\boldsymbol{x},q} \\
&\quad + \frac{\gamma}{N}\sum_{n=1}^{N} \psi_q \max\left\{\left(\frac{\sum_{m=1}^{N} \kappa_{n,m} q_n}{\sum_{m=1}^{N} \kappa_{n,m}} - q_n\right), 0\right\},
\end{aligned} \tag{3}$$

where $\langle\cdot\rangle_{\boldsymbol{x},q}$ means integration against both $\boldsymbol{x}$ and q.

Let us define

$$\rho(\boldsymbol{x}) = \frac{1}{N}\sum_{n=1}^{N} \delta(\boldsymbol{x} - \boldsymbol{x}_n)$$

and

$$m(\boldsymbol{x}) = \left\langle q, \frac{1}{N}\sum_{m=1}^{N}\delta(\boldsymbol{x}-\boldsymbol{x}_m)\delta(q-q_m)\right\rangle_{\boldsymbol{x},q} = \frac{1}{N}\sum_{m=1}^{N}\delta(\boldsymbol{x}-\boldsymbol{x}_m)q_m.$$

We have

$$\frac{1}{N}\sum_{m=1}^{N}\kappa(|\boldsymbol{x}_n-\boldsymbol{x}_m|) = \left\langle \kappa(|\boldsymbol{x}_n-\tilde{\boldsymbol{x}}|), \frac{1}{N}\sum_{m=1}^{N}\delta(\tilde{\boldsymbol{x}}-\boldsymbol{x}_m)\right\rangle_{\boldsymbol{x}} = \kappa * \rho(\boldsymbol{x}_n),$$

$$\frac{1}{N}\sum_{m=1}^{N}\kappa(|\boldsymbol{x}_n-\boldsymbol{x}_m|)q_m = \left\langle \kappa(|\boldsymbol{x}_n-\tilde{\boldsymbol{x}}|), \frac{1}{N}\sum_{m=1}^{N}\delta(\tilde{\boldsymbol{x}}-\boldsymbol{x}_m)q_m\right\rangle_{\boldsymbol{x}} = \kappa * m(\boldsymbol{x}_n),$$

where $\langle\cdot\rangle_{\boldsymbol{x}}$ means integration only in $\boldsymbol{x}$. Then, we can rewrite Eq. (3) as

$$\begin{aligned}\frac{d}{dt}\langle h^N, \psi\rangle_{\boldsymbol{x},q} = {} & \langle h^N, \psi_x v\cos\theta\rangle_{\boldsymbol{x},q} + \langle h^N, \psi_y v\sin\theta\rangle_{\boldsymbol{x},q} \\ & + \gamma\left\langle h^N, \psi_q \max\left\{\frac{\kappa * m_s}{\kappa * \rho_s} - q, 0\right\}\right\rangle_{\boldsymbol{x},q}. \end{aligned} \tag{4}$$

Via integration by parts, Eq. (4) leads to

$$h_t^N + \nabla\cdot(v(\cos\theta,\sin\theta)^T h^N) + \gamma(\max\{(q^*-q),0\}h^N)_q = 0, \tag{5}$$

where q^* is the local *average* sickness level weighted by (2):

$$q^*(t,\boldsymbol{x}) = \frac{\iint \kappa(|\boldsymbol{x}-\overline{\boldsymbol{x}}|)h(t,\overline{\boldsymbol{x}},q)q\,dq\,d\overline{\boldsymbol{x}}}{\iint \kappa(|\boldsymbol{x}-\overline{\boldsymbol{x}}|)h(t,\overline{\boldsymbol{x}},q)\,dq\,d\overline{\boldsymbol{x}}}. \tag{6}$$

Letting $k\to\infty$, we obtain the limiting kinetic model from the subsequence h^{N_k}:

$$h_t + \nabla\cdot(v(\cos\theta,\sin\theta)^T h) + \gamma(\max\{(q^*-q),0\}h)_q = 0, \tag{7}$$

where $h(t,\boldsymbol{x},q)$ is the probability of finding at time t and position $\boldsymbol{x}$ of a person with level of exposure q if $q\in[0,1)$ or a person spreading the disease if $q=1$.

It is convenient to switch to non-dimensional variables. Let D be the largest distance a pedestrian can cover in domain Ω and V_M the largest speed a person can reach. Then, we can define characteristic time $T = D/V_M$. Moreover, let ρ_M be the maximum people density per square meter. The non-dimensional quantities are: position $\hat{\boldsymbol{x}} = \boldsymbol{x}/D$, walking speed $\hat{v} = v/V_M$, time $\hat{t} = t/T$, and distribution function $\hat{h} = h/\rho_M$. For ease of notation though, we will drop the hat with the understanding that all quantities are non-dimensional unless otherwise specified.

3 Discretization in Space and Time

In this section, we present the numerical discretization of model (7).

Let us start from the space discretization. Divide the spatial domain into a number of cells $[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]\times[y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}]$ of length Δx and Δy, respectively. The discrete mesh points x_i and y_j are given by:

$$x_i = i\Delta x, \quad x_{i+1/2} = x_i + \frac{\Delta x}{2} = \left(i + \frac{1}{2}\right)\Delta x,$$
$$y_j = j\Delta y, \quad y_{j+1/2} = y_j + \frac{\Delta y}{2} = \left(j + \frac{1}{2}\right)\Delta y,$$

for $i = 0, 1, \ldots, N_x$ and $j = 0, 1, \ldots, N_y$. The contagion level domain is partitioned into subdomains $[q_{k-\frac{1}{2}}, q_{k+\frac{1}{2}}]$ of length of Δq with $k \in 1, 2, \ldots, N_q$.

$$q_k = k\Delta q, \quad q_{k+1/2} = q_k + \frac{\Delta q}{2} = \left(k + \frac{1}{2}\right)\Delta q.$$

Here, N_x, N_y, and N_q are the total number of points in $x-$, $y-$, and $q-$ directions, respectively.

Let us denote $h_{i,j,k} = h(t, x_i, y_j, q_k)$. The average level of sickness is computed using a midpoint rule for the integrals in (6):

$$q^*(t, x_i, y_j) \approx q^*_{i,j} = \frac{\sum_k \sum_{\bar{j}} \sum_{\bar{i}} \kappa_{\bar{i},\bar{j}} h_{\bar{i},\bar{j},k} q_k \Delta q \Delta x \Delta y}{\sum_k \sum_{\bar{j}} \sum_{\bar{i}} \kappa_{\bar{i},\bar{j}} h_{\bar{i},\bar{j},k} \Delta q \Delta x \Delta y}, \tag{8}$$

where if at $(x_{\bar{i}}, y_{\bar{j}})$ there is nonzero probability of finding a person who is spreading the disease:

$$\kappa_{\bar{i},\bar{j}} = \frac{R}{((x_i - x_{\bar{i}})^2 + (y_j - y_{\bar{j}})^2 + R^2)\pi},$$

otherwise $\kappa_{\bar{i},\bar{j}} = 0$.

We start with a first-order semi-discrete upwind scheme for problem (7), which reads

$$\partial_t h_{i,j,k} + \frac{\eta_{i,j,k} - \eta_{i-1,j,k}}{\Delta x} + \frac{\phi_{i,j,k} - \phi_{i,j-1,k}}{\Delta y} + \gamma \frac{\xi_{i,j,k+\frac{1}{2}} - \xi_{i,j,k-\frac{1}{2}}}{\Delta q} = 0, \tag{9}$$

where

$$\eta_{i,j,k} = v \cos\theta \, h_{i,j,k}, \quad \phi_{i,j,k} = v \sin\theta \, h_{i,j,k}, \tag{10}$$

$$\xi_{i,j,k+\frac{1}{2}} = \max\left\{q^*_{i,j} - q_{k+\frac{1}{2}}, 0\right\} h_{i,j,k}. \tag{11}$$

Next, let us discretize in time. Let $\Delta t \in \mathbb{R}$, $t^n = t_0 + n\Delta t$, with $n = 0, \ldots, N_T$ and $t_f = t_0 + N_T \Delta t$, where t_f is the end of the time interval under consideration. Moreover, we denote by y^n the approximation of a generic quantity y at the time t^n. For the time discretization of problem (9), we use the forward Euler scheme:

$$h^{l+1}_{i,j,k} = h^l_{i,j,k} - \Delta t \left(\frac{\eta^l_{i,j,k} - \eta^l_{i-1,j,k}}{\Delta x} + \frac{\phi^l_{i,j,k} - \phi^l_{i,j-1,k}}{\Delta y} + \gamma \frac{\xi^l_{i,j,k+\frac{1}{2}} - \xi^l_{i,j,k-\frac{1}{2}}}{\Delta q} \right). \tag{12}$$

To construct a scheme that is of second order, we add a flux limiter. Let φ be a slope limiter function. For example, one could choose the Van Leer function:

$$\varphi(\theta) = \frac{|\theta| + \theta}{1 + |\theta|}. \tag{13}$$

The space discretized Eq. (7) now reads

$$\begin{aligned} \partial_t h_{i,j,k} + \frac{\overline{\eta}_{i,j,k} - \overline{\eta}_{i-1,j,k}}{\Delta x} + \frac{\overline{\phi}_{i,j,k} - \overline{\phi}_{i,j-1,k}}{\Delta y} \\ + \gamma \frac{\xi_{i,j,k+\frac{1}{2}} - \xi_{i,j,k-\frac{1}{2}}}{\Delta q} + \gamma \frac{C_{i,j,k+\frac{1}{2}} - C_{i,j,k-\frac{1}{2}}}{\Delta q} = 0, \end{aligned} \tag{14}$$

where

$$\overline{\eta}_{i,j,k} = \eta_{i,j,k} + \frac{\eta_{i+1,j,k} - \eta_{i,j,k}}{2} \varphi\left(\frac{\eta_{i,j,k} - \eta_{i-1,j,k}}{\eta_{i+1,j,k} - \eta_{i,j,k}} \right), \tag{15}$$

$$\overline{\phi}_{i,j,k} = \phi_{i,j,k} + \frac{\phi_{i,j+1,k} - \phi_{i,j,k}}{2} \varphi\left(\frac{\phi_{i,j,k} - \phi_{i,j-1,k}}{\phi_{i,j+1,k} - \phi_{i,j,k}} \right), \tag{16}$$

$$q_{k+\frac{1}{2}} = \frac{q_k + q_{k+1}}{2}, \; s_{i,j,k+\frac{1}{2}} = q^*_{i,j} - q_{k+\frac{1}{2}}$$

$$W_{i,j,k-\frac{1}{2}} = h_{i,j,k} - h_{i,j,k-\frac{1}{2}}$$

$$C_{i,j,k+\frac{1}{2}} = \frac{1}{2} \left| s_{i,j,k+\frac{1}{2}} \right| \left(1 - \frac{\Delta t}{\Delta q} \left| s_{i,j,k+\frac{1}{2}} \right| \right) W_{i,j,k-\frac{1}{2}} \varphi\left(\frac{W_{i,j,\mathbf{k}-\frac{1}{2}}}{W_{i,j,k-\frac{1}{2}}} \right). \tag{17}$$

In Eq. (17), the subscript **k** is k-1 if $s_{i,j,k-\frac{1}{2}} > 0$ and k+1 if $s_{i,j,k-\frac{1}{2}} < 0$. Recall that $\eta_{i,j,k}$ in (15) and $\phi_{i,j,k}$ in (16) are defined in (10), while $\xi_{i,j,k+\frac{1}{2}}$ is defined in (11). Thanks to the flux limiter, scheme (14) is of second-order scheme in velocity.

For the time discretization of problem (14), we adopt again the forward Euler scheme

$$h_{i,j,k}^{l+1} = h_{i,j,k}^{l} - \Delta t\Big(\frac{\overline{\eta}_{i,j,k}^{l} - \overline{\eta}_{i-1,j,k}^{l}}{\Delta x} + \frac{\overline{\phi}_{i,j,k}^{l} - \overline{\phi}_{i,j-1,k}^{l}}{\Delta y} + \gamma \frac{\xi_{i,j,k+\frac{1}{2}}^{l} - \xi_{i,j,k-\frac{1}{2}}^{l}}{\Delta q} + \gamma \frac{C_{i,j,k+\frac{1}{2}}^{l} - C_{i,j,k-\frac{1}{2}}^{l}}{\Delta q}\Big). \tag{18}$$

The time step Δt for scheme (12) or (18) is set as

$$\Delta t = \frac{1}{2} \min\left\{ \frac{\Delta x}{\max_k q_k}, \frac{\Delta y}{\max_k q_k}, \frac{\Delta q}{2\gamma \max_k q_k} \right\} \tag{19}$$

to satisfy the Courant–Friedrichs–Lewy (CFL) condition.

4 Numerical Results

We assess the approach presented in Sect. 3 through a series of tests. For all the problems, the computational domain in the xy-plane is $[0, 10] \times [0, 10]$ m^2 before non-dimensionalization, while $q \in [0, 1]$. The dimensionless quantities are obtained by using the following reference quantities: $D = 10\sqrt{2}$ m, $V_M = 1$ m/s, $T = 10\sqrt{2}$ s, and $\rho_M = 7$ people/m^2. In all the tests, we take the initial density to be constant in space and equal to ρ_M. So, the group occupying the walkable domain consists of 700 people.

We set the interaction distance $R = 1$ m since this choice makes the value of the kernel function relatively small at a distance of 2 m (or about 6 ft). See Fig. 1.

To discretize the walkable domain, we take $\Delta x = \Delta y = 0.1$ m, while we set $\Delta q = 0.01$. We will consider two values for the contagion strength: $\gamma = 50$ and $\gamma = 10$. The time step is set to $\Delta t = 5 \cdot 10^{-5}$ s for $\gamma = 50$ and to $\Delta t = 25 \cdot 10^{-5}$ s for $\gamma = 10$ according to (19). First, we keep the group of people still (i.e., $v = 0$) to make sure that the level of exposure evolves as expected. Then, in a second set of tests, we let people walk and observe how the motion affects the spreading of the disease. We run each simulation for $t \in (0, 10]$ s.

4.1 Tests with $v = 0$

We adapt two initial conditions from [29]:

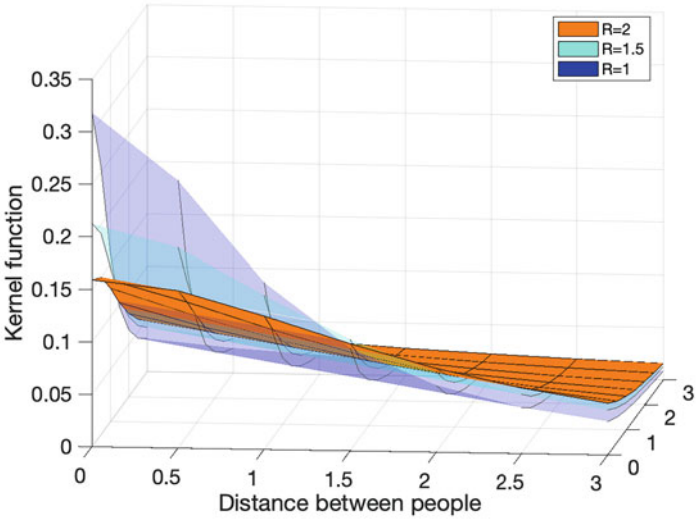


Fig. 1 Kernel function vs. the distance between people for interaction radius $R = 1, 1.5, 2$

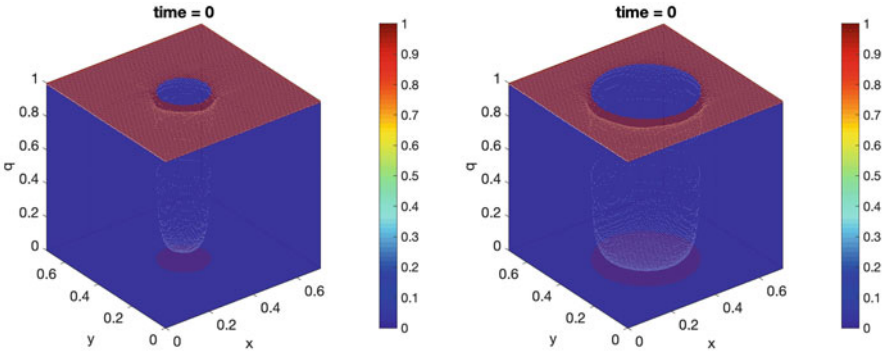


Fig. 2 Initial conditions IC1 (left) and IC2 (right)

– IC1: people who are certainly spreading (i.e., $q = 1$) are located outside the circle of radius 1 m centered at (5, 5) m, while within that circle we place people who have certainly not been exposed (i.e., $q = 0$). See Fig. 2 (left).
– IC2: people who are certainly spreading (i.e., $q = 1$) are located outside the circle of radius 2 m centered at (5, 5) m, while the rest of the people located within that circle have certainly not been exposed (i.e., $q = 0$). See Fig. 2 (right).

The difference between the two boundary conditions is that all the healthy people in IC1 are exposed to spreading people, while some healthy people in IC2 are not exposed to spreading people. Thus, we expect that the level of exposure for the people who are centrally located rises faster for IC1 than for IC2.

Figure 3 shows the evolution of the distribution density h for initial conditions IC1 and IC2 with $\gamma = 50$, together with a view of the results on a section of the 3D domain. We observe that the level of exposure of the central group of healthy people in IC1 increases quickly. In particular, we see that the shape from q^* quickly moves away from being close to a sharp discontinuity and gets closer to a paraboloid, since the level of exposure increases faster for the people closer to the circle separating healthy people from spreading people. As expected, the rise in the level of exposure is much slower for the simulation with initial condition IC2. This is evident when comparing the second and fourth row in Fig. 3. Moreover, notice that the healthy people close to the center of the circle in IC2 get very little exposure to spreading people.

Next, we consider initial condition IC1 and compare the evolution of the distribution density h for $\gamma = 50$ and $\gamma = 10$. See Fig. 4. We see that the level of exposure of the central group of healthy people increases faster for larger values of γ, as expected. Parameter γ plays a key role in the evolution of the disease spreading according to model (7). If one was interested in modeling a realistic scenario, such as the spreading of COVID-19 in a ER waiting room, γ would have to be carefully tuned.

This first set of tests was meant to verify our implementation of method described in Sect. 3 and to check that the disease spreading term in Eq. (7) (i.e., the third term on the left-hand side) produced the expected outcomes. Next, we will set people in motion.

4.2 Tests with Prescribed Walking Velocity

We assign to all people walking speed $v = 1$ m/s and walking direction $\theta = \pi/4$, as if they were headed to the upper right corner of the domain in the xy-plane. Once the people in the spreading phase of the disease have left the domain, we assume they cannot spread to the people in the walkable domain anymore.

We consider again IC1 and IC2 with $\gamma = 50$. Figure 5 shows the evolution of the distribution density h for both initial conditions. As one would expect, the motion contributes to lowering the exposure level in the both cases, since some of the spreading people leave the domain as soon as the motion starts. This difference is obvious when one compares the second (resp., fourth) row in Fig. 3 with the second (resp., fourth) row in Fig. 5.

Finally, we modify the initial conditions to show that our model can handle also more complex scenarios featuring uncertainties. IC1 and IC2 are changed to:

– IC1-bis: people are positioned like in IC1 but the probabilities of finding people with $q = 1$ and $q = 0$ is reduced and another value of q is assigned at a given (x, y), i.e., at every location in the walkable domain the initial distribution has two bumps in q. See Fig. 6 (left).

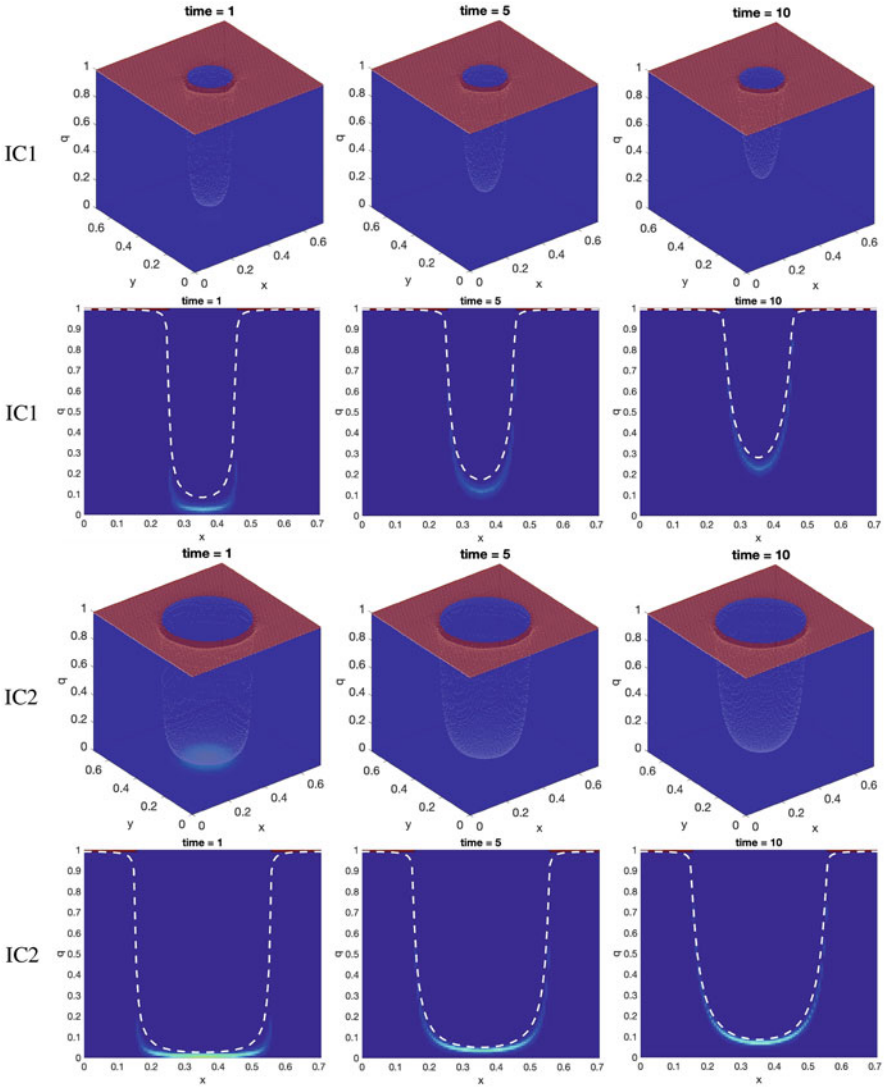


Fig. 3 Tests with $v = 0$: evolution of the distribution density h for initial condition IC1 (first row) and corresponding results on section $y = 0.5$ (second row), and evolution of h for initial condition IC2 (third row) and corresponding results on section $y = 0.5$ (fourth row). Note that in all the subfigures time is dimensional while space is non-dimensional. In both cases, we set $\gamma = 50$. The white dashed line in the images on row two and four represents q^*. The legend for all the images is the same as in Fig. 2

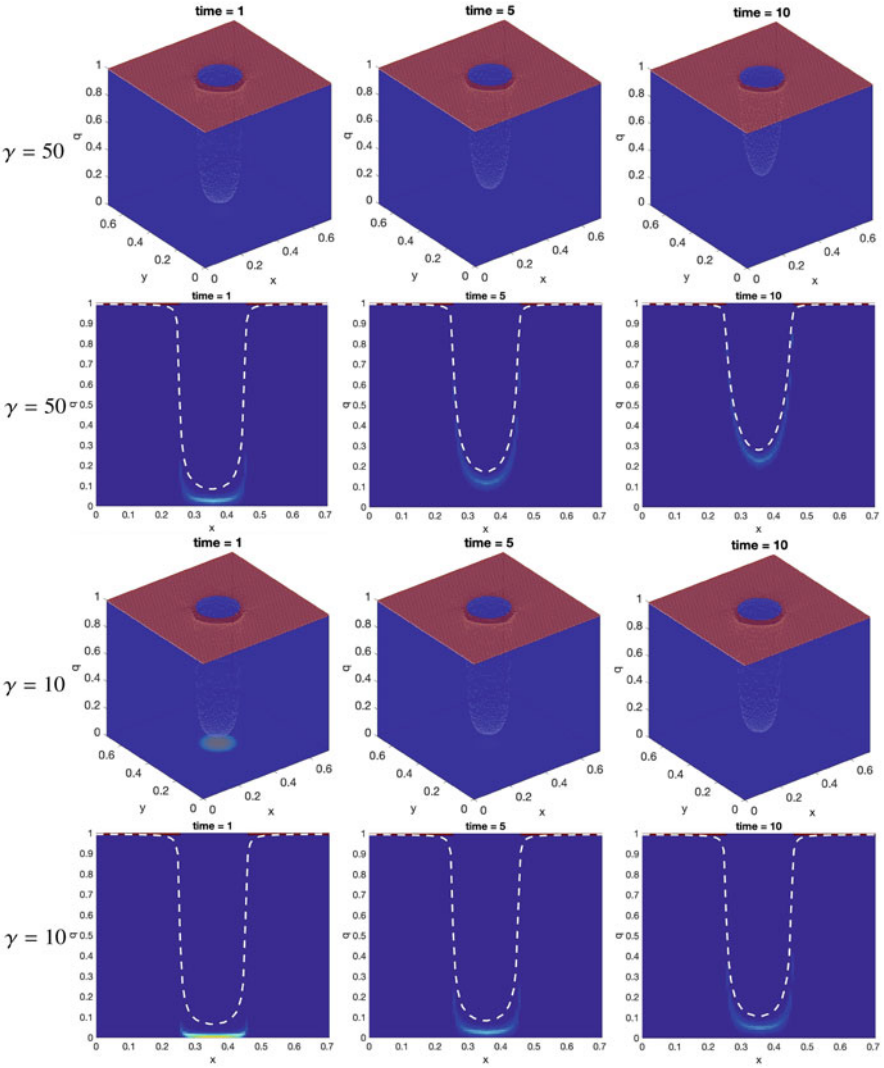


Fig. 4 Tests with $v = 0$: evolution of the distribution density h for $\gamma = 50$ (first row) and corresponding results on section $y = 0.5$ (second row), and evolution of h for $\gamma = 10$ (third row) and corresponding results on section $y = 0.5$ (fourth row). In both cases, the initial condition is IC1. Note that in all the subfigures time is dimensional while space is non-dimensional. The white dashed line in the images on row two and four represents q^*. The legend for all the images is the same as in Fig. 2

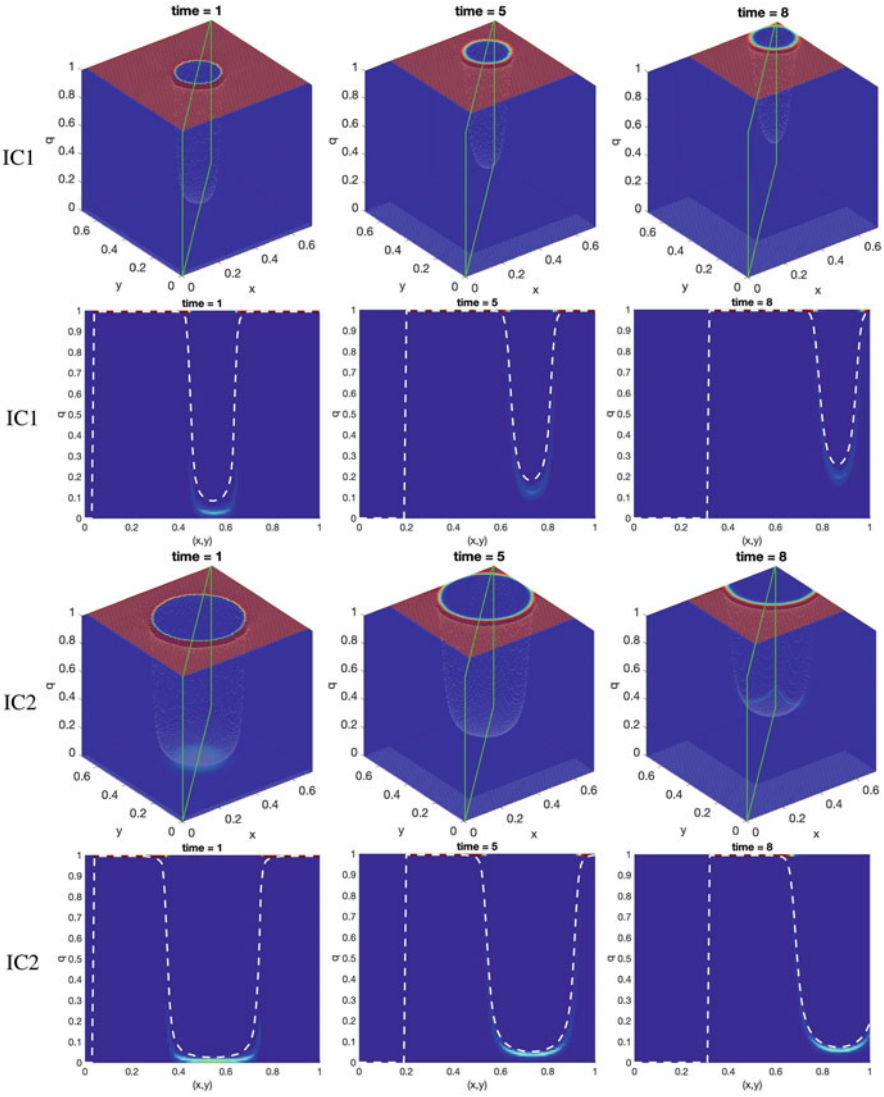


Fig. 5 Tests with prescribed walking velocity: the first row shows the evolution of the distribution density h for initial condition IC1 and highlighted in green is the section whose results are reported in the second row. The third row shows the evolution of h for initial condition IC2 and highlighted in green is the section whose results are reported in the fourth row. Note that in all the subfigures time is dimensional while space is non-dimensional. For both cases, $\gamma = 50$. The white dashed line in the images on row two and four represents q^*. The legend for all the images is the same as in Fig. 2

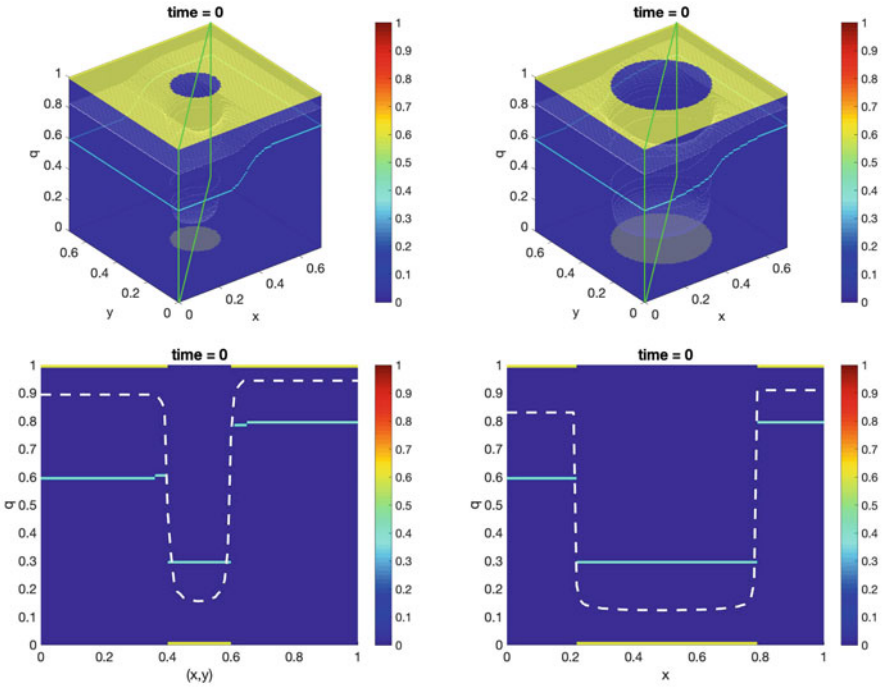


Fig. 6 Tests with prescribed walking velocity: initial conditions IC1-bis (top left) and IC2-bis (top right). The initial data on the section highlighted in green in the top panels is reported in the corresponding bottom panel

– IC2-bis: people are positioned like in IC2 but the probabilities of finding people with $q = 1$ and $q = 0$ is reduced and, similarly to IC1-bis, another value of q is assigned at a given (x, y). See Fig. 6 (right).

Figure 7 shows the evolution of the distribution density h for initial conditions IC1-bis and IC2-bis with $\gamma = 50$.

5 A More Complex 2D Kinetic Model

In this section, we remove the simplifying assumption used in Sect. 2, i.e., walking speed v and direction θ are no longer assumed to be given. Instead, they result from the interaction with the surrounding people. Following [1, 27, 29], we account for the granular feature of the system (i.e., the fact that the distance between pedestrians can range from small to large) by discretizing θ. Instead of being continuous, variable θ can only take values in the set:

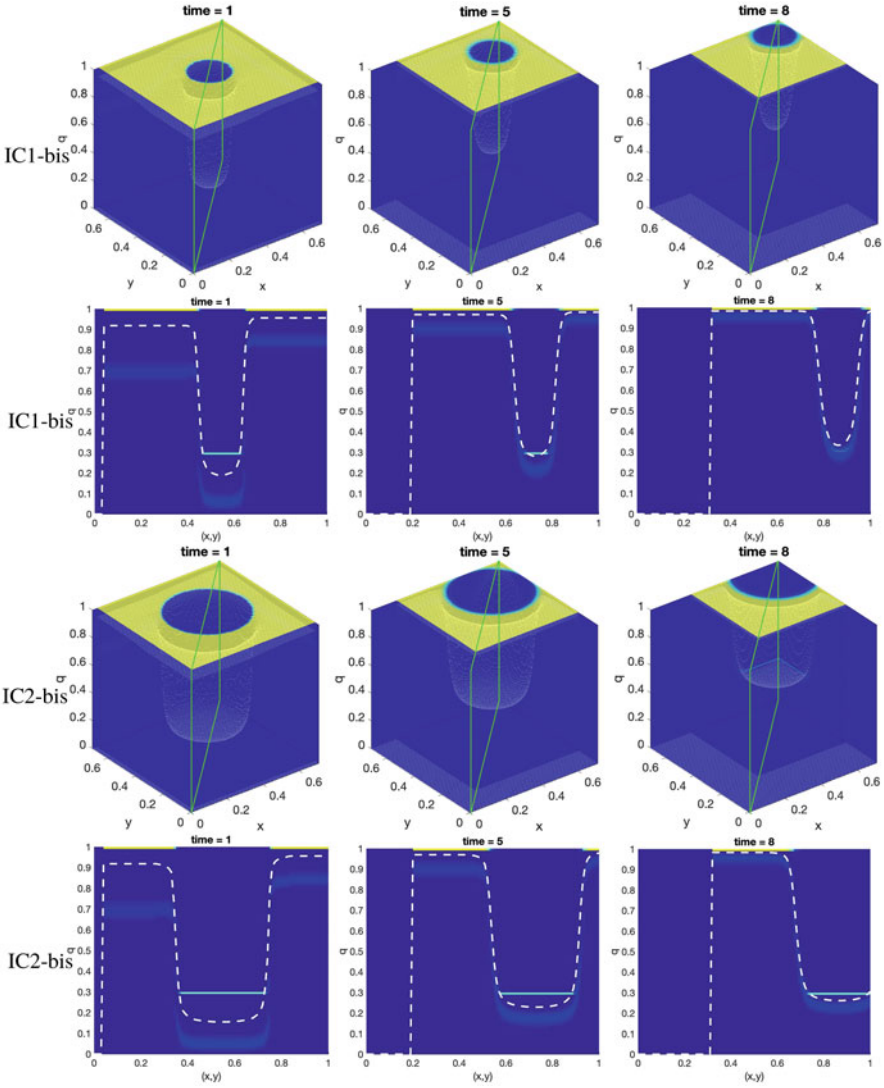


Fig. 7 Tests with prescribed walking velocity: the first row shows the evolution of the distribution density h for initial condition IC1-bis and highlighted in green the section whose results are reported in the second row. The third row shows the evolution of h for initial condition IC2-bis and highlighted in green the section whose results are reported in the fourth row. Note that in all the subfigures time is dimensional while space is non-dimensional. In both cases, we set $\gamma = 50$. The white dashed line in the images on row two and four represents q^*. The legend for all the images is the same as in Fig. 6

$$I_\theta = \left\{ \theta_i = \frac{i-1}{N_d} 2\pi : i = 1, \ldots, N_d \right\},$$

where N_d is the maximum number of possible directions. As for the walking speed v, we assume it depends on the local level of congestion as demonstrated by experimental studies. Thus, variable v is treated as a deterministic variable.

Let $\boldsymbol{x} = (x, y)$ denote position. For the above-mentioned assumptions on θ and v, the distribution function can be written as

$$f(t, \boldsymbol{x}, \theta) = \sum_{i=1}^{N_d} f^i(t, \boldsymbol{x}) \delta(\theta - \theta_i), \tag{20}$$

where $f^i(t, \boldsymbol{x})$ represents the people who, at time t and position $\boldsymbol{x}$, move with direction θ_i, and δ denotes the Dirac delta function. Like for the model in Sect. 2, it is convenient to switch to non-dimensional variables: position $\hat{\boldsymbol{x}} = \boldsymbol{x}/D$, walking speed $\hat{v} = v/V_M$, time $\hat{t} = t/T$, and distribution function $\hat{f} = f/\rho_M$. Once again, we will drop the hat for ease of notation though, with the understanding that all variables from now on are non-dimensional. Due to the normalization of f and the f_i functions, the dimensionless local density is obtained by summing the distribution functions over the set of directions:

$$\rho(t, \boldsymbol{x}) = \sum_{i=1}^{N_d} f^i(t, \boldsymbol{x}). \tag{21}$$

The model in [1, 27, 29] features two key parameters:

- $\alpha \in [0, 1]$, which represents the quality of the walkable domain: if $\alpha = 1$ the domain is clear, while if $\alpha = 0$ an obstruction is present. Note that α could be dependent on space (and possibly time) in a prescribed way.
- $\varepsilon \in [0, 1]$, which represents the overall level of stress: $\varepsilon = 0$ indicates no stress, while $\varepsilon = 1$ indicates a high stress situation.

Parameter α plays a role in determining the walking speed: where $\alpha = 1$ people can walk at the desired speed (i.e., V_M), while where $\alpha = 0$ people are forced to slow down or stop. See, e.g., [27] on how to define the walking speed $v = v[\rho](t, \boldsymbol{x}, \alpha)$, with the square brackets denoting that v depends on ρ in a functional way. Parameter ε plays a role in the selection of the walking direction when interacting with other pedestrians. Indeed, the choice of the walking direction among the possible directions in I_θ results from four competing factors:

1. The goal to reach a target, e.g., the gate of an airport terminal
2. The desire to avoid collisions with the walls and other obstacles
3. The tendency to look for less crowded areas
4. The tendency to follow the stream or "herd"

Parameter ε weighs between 3 (prevailing when stress is low) and 4 (prevailing when stress is high).

In order to explain how the four factors are modeled, we need to introduce some terminology. Interactions involve three types of active particles (or people):

- *Test particle* with state $(\boldsymbol{x}, \theta_i)$: they are representative of the whole system.
- *Candidate particle* with state $(\boldsymbol{x}, \theta_h)$: they can reach in probability the state of the test particle after individual-based interactions with the venue (walls, obstacles, or target) or with field particles.
- *Field particle* with state $(\boldsymbol{x}, \theta_k)$: their interactions with candidate particles determines a possible change of state.

Note that if a test particle changes their state (in probability) into that of the test of the test particle as a result of the interactions with field people, the test particle loses their state.

We call $\mathcal{A}_h(i)$ the probability that a candidate particle with direction θ_h adjusts their direction to θ_i (the direction of the test particle) due to the interaction with the venue. The transition probability that the candidate particle changes their direction to θ_i (direction of the test particle) in the search for less crowded areas if their stress level is low or to θ_k (direction of the field particle) if their stress level is high is denoted by $\mathcal{B}_{hk}(i)$. The sets of all transition probabilities $\mathcal{A} = \{\mathcal{A}_h(i)\}_{h,i=1,\dots,N_d}$ and $\mathcal{B} = \{\mathcal{B}_{hk}(i)\}_{h,k,i=1,\dots,N_d}$ form the so-called tables of games that model the game played by active particles interacting with the venue and other particles, respectively. See, e.g., [27] for more details.

In order to state the mathematical model, we need two more ingredients:

- The *interaction rate* with the venue $\mu[\rho]$: it models the frequency of interactions between candidate particles and the venue and it is assumed to have a functional dependence on ρ. In fact, it is easier for pedestrians to see the walls, possible obstacles, and their target if density is low.
- The *interaction rate* with other pedestrians $\eta[\rho]$: it defines the number of binary encounters per unit time. This rate depends on ρ in a functional way too since if the local density increases the interaction rate also increases.

For possible ways to set $\mu[\rho]$ and $\eta[\rho]$, see, e.g., [27].

In order to derive the mathematical model, one has to take the balance of particles in an elementary volume of the space of microscopic states, considering the net flow into such volume due to transport and interactions. Let $\mathcal{J}^i[f]$ be the net balance of particles that move with direction θ_i:

$$\mathcal{J}^i[f](t,\boldsymbol{x},q) = \mu[\rho]\left(\sum_{h=1}^{N_d}\int_0^1 \mathcal{A}_h(i)\,f^h(t,\boldsymbol{x},q_h)dq_h - f^i(t,\boldsymbol{x},q)\right)$$

$$+\,\eta[\rho]\left(\sum_{h,k=1}^{N_d}\int_0^1\int_0^1 \mathcal{B}_{hk}(i)\,f^h(t,\boldsymbol{x},q_h)\,f^k(t,\boldsymbol{x},q_k)dq_h dq_k\right.$$

$$-f^i(t,\boldsymbol{x},q)\rho(t,\boldsymbol{x})\Big) \tag{22}$$

due to interactions with the venue (first term at the left-hand side of (22)) and with the surrounding particles (second term at the left-hand side of (22)). Then, the model is given by:

$$\begin{aligned}\frac{\partial f^i}{\partial t} + \nabla\cdot\Big(v(\cos\theta_i,\sin\theta_i)^T f^i(t,\boldsymbol{x},q)\Big)\\ = \mathcal{J}^i[f](t,\boldsymbol{x},q) + \gamma\frac{\partial(\max\{(q^*-q),0\}f^i)}{\partial q}\end{aligned} \tag{23}$$

for $i = 1, 2, \ldots, N_d$. Problem (22)–(23) is completed with Eq. (21) for the density and a suitable law $v = v[\rho](t,\boldsymbol{x},\alpha)$ that relates the walking speed to the density.

While modeling motion in 2 dimensions (x and y), Eq. (23) is a 3D problem in variables x, y, and q. Its numerical approximation requires a carefully designed scheme to contain the computational cost. Here, we are only going to present ideas on how to design such a scheme. Numerical results will be presented in a follow-up paper. Following what we have done in [30] for emotional contagion, we will split model (23) into subproblems that are easier to solve using operator splitting (see, e.g., [22]). For example, with the Lie splitting algorithm one could break problem (23) into: (i) a pure advection problem that features the contagion term, i.e., the last term in Eq. (23), and (ii) a problem involving the interaction with the venue and other pedestrians. Then, suitable finite difference scheme will be applied for the space and time discretization of problems (i) and (ii).

6 Conclusions

We presented a way to design a kinetic type model to simulate the early stage of an infectious disease spreading in an intermediate size population occupying a confined environment for a short period of time.

In order to focus on the spreading mechanism, in the first part of the chapter we adopted a strong simplifying assumption: people's walking speed and direction are given. The disease spreading is modeled using three main ingredients: an additional variable for the level of exposure to people spreading the disease, a parameter for the contagion interaction strength, and a kernel function that is a decreasing function of the distance between a person and a spreading individual. For such a simplified model, we propose a first-order and a second-order finite difference scheme. We tested our numerical approach on simple 2D problems and verified that the level of exposure evolved as expected, even in scenarios featuring uncertainties.

In the second part of the chapter, we used ideas from the simplified model to incorporate disease spreading in a kinetic theory approach for crowd dynamics. The appealing feature of kinetic (or mesoscopic) models for crowd dynamics is the

flexibility in accounting for multiple interactions (hard to achieve in microscopic models) and heterogeneous behavior in people (hard to achieve in macroscopic models). Modeling crowd dynamics and disease contagion in two spatial dimensions requires the solution of a 3D problem, the additional variable being the contagion level. The design of an efficient numerical scheme for such problem will be an object of a follow-up paper.

Acknowledgments This work has been partially supported by NSF through grant DMS-1620384.

References

1. J.P. Agnelli, F. Colasuonno, D. Knopoff, A kinetic theory approach to the dynamics of crowd evacuation from bounded domains. Math. Models Methods Appl. Sci. **25**(01), 109–129 (2015)
2. V.V. Aristov, Biological systems as nonequilibrium structures described by kinetic methods. Results Phys. **13**, 102232 (2019)
3. M. Asano, T Iryo, M. Kuwahara, Microscopic pedestrian simulation model combined with a tactical model for route choice behaviour. Transp. Res. C Emerg. Technol. **18**(6), 842–855 (2010)
4. B. Aylaj, N. Bellomo, L. Gibelli, A. Reali, A unified multiscale vision of behavioral crowds. Math. Models Methods Appl. Sci. **30**(1), 1–22 (2020)
5. N. Bellomo, A. Bellouquid, On the modeling of crowd dynamics: looking at the beautiful shapes of swarms. Netw. Heterogeneous Media **6**(3), 383–399 (2011)
6. N. Bellomo, L. Gibelli, Toward a mathematical theory of behavioral-social dynamics for pedestrian crowds. Math. Models Methods Appl. Sci. **25**(13), 2417–2437 (2015)
7. N. Bellomo, L. Gibelli, Behavioral crowds: modeling and Monte Carlo simulations toward validation. Comput. Fluids **141**, 13–21 (2016)
8. N. Bellomo, A. Bellouquid, D. Knopoff, From the microscale to collective crowd dynamics. SIAM Multiscale Model. Simul. **11**(3), 943–963 (2013)
9. N. Bellomo, A. Bellouquid, L. Gibelli, N Outada, A quest towards a mathematical theory of living systems, in *Modeling and Simulation in Science, Engineering and Technology* (Birkhäuser, Basel, 2017)
10. N. Bellomo, L. Gibelli, N. Outada, On the interplay between behavioral dynamics and social interactions in human crowds. Kinet. Relat. Models **12**(2), 397–409 (2019)
11. N. Bellomo, R. Bingham, M.K. Chaplain, G. Dosi, G. Forni, D.A. Knopoff, J. Lowengrub, R. Twarock, M.E. Virgillito, A multiscale model of virus pandemic: heterogeneous interactive entities in a globally connected world. Math. Models Methods Appl. Sci. **30**(8), 1591–1651 (2020)
12. N. Bellomo, D. Burini, G. Dosi, L. Gibelli, D.A. Knopoff, N. Outada, P. Terna, M.E. Virgillito, What is life? A perspective of the mathematical kinetic theory of active particles. Math. Models Methods Appl. Sci. **31**(9), 1821–1866 (2021)
13. A.L. Bertozzi, J. Rosado, M.B. Short, L. Wang, Contagion shocks in one dimension. J. Stat. Phys. **158**, 647–664 (2015)
14. P. Billingsley, *Convergence of Probability Measures* (Wiley, Hoboken, 1999)
15. W. Boscheri, G. Dimarco, L. Pareschi, Modeling and simulating the spatial spread of an epidemic through multiscale kinetic transport equations. Math. Models Methods Appl. Sci. **31**, 1–39 (2021)
16. M. Chraibi, A. Tordeux, A. Schadschneider, A. Seyfried, *Modelling of Pedestrian and Evacuation Dynamics* (Springer, New York, 2019), pp. 649–669

17. E. Cristiani, B. Piccoli, A. Tosin, *Multiscale Modeling of Pedestrian Dynamics* (Springer, Cham, 2014)
18. J. Dai, X. Li, L. Liu, Simulation of pedestrian counter flow through bottlenecks by using an agent-based model. Phys. A Stat. Mech. Appl. **392**(9), 2202–2211 (2013)
19. P. Derjany, S. Namilae, D. Liu, A. Srinivasan, Multiscale model for the optimal design of pedestrian queues to mitigate infectious disease spread. PLOS ONE **15**, 1–21 (2020)
20. F. Flandoli, E. La Fauci, M. Riva, Individual-based Markov model of virus diffusion: comparison with Covid-19 incubation period, serial interval and regional time series. Math. Models Methods Appl. Sci. **31**, 907–939 (2021)
21. M. Gatto, E. Bertuzzo, L. Mari, S. Miccoli, L. Carraro, R. Casagrandi, A. Rinaldo, Spread and dynamics of the covid-19 epidemic in Italy: effects of emergency containment measures. Proc. Natl. Acad. Sci. **117**(19), 10484–10491 (2020)
22. R. Glowinski, Finite element methods for incompressible viscous flow, in *Handbook of Numerical Analysis*, ed. by P.G. Ciarlet, J.L. Lions, vol. 9 (North-Holland, Amsterdam, 2003)
23. T. Harweg, D. Bachmann, F. Weichert, Agent-based simulation of pedestrian dynamics for exposure time estimation in epidemic risk assessment. J Public Health (Berl.) (2021). https://doi.org/10.1007/s10389-021-01489-y
24. D. Helbing, P. Molnár, Social force model for pedestrian dynamics. Phys. Rev. E **51**, 4282–4286 (1998)
25. H.W. Hethcote, The mathematics of infectious diseases. SIAM Rev. **42**(4), 599–653 (2000)
26. R.L. Hughes, The flow of human crowds. Annu. Rev. Fluid Mech. **35**(1), 169–182 (2003)
27. D. Kim, A. Quaini, A kinetic theory approach to model pedestrian dynamics in bounded domains with obstacles. Kinet. Relat. Models **12**(6), 1273–1296 (2019)
28. D. Kim, A. Quaini, Coupling kinetic theory approaches for pedestrian dynamics and disease contagion in a confined environment. Math. Models Methods Appl. Sci. **30**(10), 1893–1915 (2020)
29. D. Kim, A. Quaini, A kinetic theory approach to model crowd dynamics with disease contagion (2021). arXiv:2103.15151
30. D. Kim, K. O'Connell, W. Ott, A. Quaini, A kinetic theory approach for 2D crowd dynamics with emotional contagion. Math. Models Methods Appl. Sci. **31**(06), 1–26 (2021). https://arxiv.org/abs/2012.08108
31. S.C. Mpeshe, N. Nyerere, Modeling the dynamics of coronavirus disease pandemic coupled with fear epidemics. Computat. Math. Methods Med. **2021**, 6647425 (2021)
32. K. Rathinakumar, A. Quaini, A microscopic approach to study the onset of a highly infectious disease spreading. Math. Biosci. **329**, 108475 (2020)
33. A. Schadschneider, A. Seyfried, Empirical results for pedestrian dynamics and their implications for modeling. Netw. Heterogeneous Media **6**(3), 545–560 (2011)
34. A. Shende, M.P. Singh, P. Kachroo, Optimization-based feedback control for pedestrian evacuation from an exit corridor. IEEE Trans. Intell. Transp. Syst. **12**(4), 1167–1176 (2011)
35. A.U.K Wagoum, A. Tordeux, W. Liao, Understanding human queuing behaviour at exits: an empirical study. R. Soc. Open Sci. **4**(1), 160896 (2017)
36. L. Wang, M.B. Short, A.L. Bertozzi, Efficient numerical methods for multiscale crowd dynamics with emotional contagion. Math. Models Methods Appl. Sci. **27**(1), 205–230 (2017)

Multiscale Derivation of a Time-Dependent SEIRD Reaction–Diffusion System for COVID-19

Mohamed Zagour

1 Introduction

The outbreak of the new coronavirus, called COVID-19, caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) appeared in December 2019, apparently occurred in Wuhan, China. The spread of the epidemics has been very fast, thus covering almost all countries in the world. The World Health Organization declared it to be a pandemic on March 11, 2020. Thus, pandemic has severely affected the economy, health, and security of the society all over the world. Data are impressive, as by February 2021 more than 100 million people have been infected and more than 2 million people died [1].

Mathematical models can contribute both to crisis managers and to research in epidemiology, however without naively claiming that mathematics can tackle the problem of derivation of models by a standing alone approach. For example, models can depict a variety of epidemic scenarios depending on the models' parameters. In addition, they can contribute to a deeper understanding of the contagion mechanisms. Accordingly, models may help in decision making, for instance, about containment measures, lock-down, and vaccination campaigns.

Several models have been proposed to describe the evolution of epidemics, which can be denoted as collective or network models. Collective models describe the spread of the epidemic in a population using a limited number of collective variables with a small number of parameters. For instance, logistic models [27, 41], generalized growth models [16], Richards models [34], susceptible-infected-recovered (SIR) models [4, 27], and SEIR (susceptible-exposed-infectious-removed) models [16].

M. Zagour (✉)
Euromed Research Center, INSA Euro-Mediterranean, Euromed University of Fes, Fès, Morocco
e-mail: m.zagour@insa.ueuromed.org

N. Bellomo, M. A. J. Chaplain (eds.), *Predicting Pandemics in a Globally Connected World, Volume 1*, Modeling and Simulation in Science, Engineering and Technology, https://doi.org/10.1007/978-3-030-96562-4_10

Note that SIR, SEIR, and other similar models belong to the class of compartmental models, see [13, 16, 35, 37], while an exhaustive presentation, which includes qualitative analysis and biological applications, is delivered in [32].

Network models treat a population as a network of interacting individuals, and the contagion process is described at the microscopic level; see [24, 42]. Data on the spread of the epidemics are available in [19, 29, 31]. Modeling of vaccination dynamics and medical actions are treated in [28, 39].

New concepts that look forward to challenging research perspectives are given in [18], where the author introduces the interested reader to conceivable derivation of compartmental models as natural development of classical SEIR models but also new concepts are proposed on the contagion dynamics on short- and long-range interactions somehow related to droplets and breathing, on the role of transportation networks, on the immune reaction to the virus, and on the role of vaccine. See also [20] for developments of SEIR models to account for Covid-19 biological dynamics. A multiscale approach is proposed in [8] by coupling the individual-based contagion dynamics to that of the in-host competition in the lung of infected individuals by a dynamics, where full recovery, need of hospitalization, and eventual death are related to proliferation of the virus contrasted by the virus. The interested reader is addressed to [8, 18] not only for a broad reference to the existing literature but also for various challenging research perspectives.

In this chapter, we deal with a multiscale approach to derive time-dependent SEIRD reaction–diffusion system (1) from kinetic theory model by using the micro–macro decomposition method. First, we rewrite the kinetic model as a coupled system of microscopic part and macroscopic one and subsequently macroscopic models are derived by low order asymptotic expansions in terms of a small parameter. This approach has been applied to the micro–macro application in different fields. For instance, chemotaxis phenomena related to the celebrated Keller–Segel model [5] and formation of patterns induced by cross-diffusion in a fluid [3, 6, 9]. This technique motivated the design of numerical tools that preserve the asymptotic property, for instance [23, 26]. These methods design stable numerical schemes in the limit along the transition from kinetic to macroscopic regimes.

Section 2 presents a phenomenological derivation of a macro-scale model of virus contagion and diffusion in space. In Sect. 3 we briefly present the micro–macro method, which leads to the derivation of system (1) from a kinetic model. Section 4 is devoted to the development of an asymptotic preserving numerical scheme in one dimension, inspired by paper [3]. The method guarantees uniform stability with respect to the parameter ε, related to the mean distance between particles, as well as consistency with the reaction–diffusion limit. In Sect. 5, we provide some numerical simulations obtained with micro–macro formulation and with the macroscopic scheme, where we show the asymptotic preserving numerical scheme property. In addition, we demonstrate the effect of the diffusion terms in system (1), and its sensitivity with respect to the transmission function.

2 Phenomenological Modeling of Diffusion Population Dynamics

Let us consider a population constituted by N_0 individuals, which can be subdivided into a number of sub-population, in short SP, each characterized by a different biological state. In more details we consider the following SP whose states are defined by their number, referred to as N_0, depending on time and space, where individuals correspond to:

1. $N(t, x)$ Alive
2. $S(t, x)$ Susceptible
3. $E(t, x)$ Exposed
4. $I(t, x)$ Infected
5. $R(t, x)$ Recovered
6. $D(t, x)$ Died

The aforementioned normalization with respect to N_0 implies that

$$N(t,x)+S(t,x)+E(t,x)+I(t,x)+R(t,x)+D(t,x)=1, \qquad t \geq 0, \qquad x \in \Omega,$$

where Ω is a bounded domain within which the population is confined.

The derivation of the mathematical model can be obtained according to the following assumptions:

1. Individuals diffuse within Ω by a diffusion coefficient $\varphi(x)$ depending on a spatial distribution, which considers the preferred directions of propagation.
2. The interaction dynamics is modeled by a source term involving the different interactions of SP as shown in Fig. 1.
3. Modeling of interactions' accounts uses the parameters reported in Table 1, which also reports the parameters underlying the assumptions of the models of the interactions shown in Fig. 1.
4. Susceptible may become Exposed with a transmission function $\beta(t)$ due to contact with infectious individuals. While exposed can become infectious with a ξ rate and infectious SP recover with a γ rate.
5. A time-dependent transmission rate function is considered. It incorporates the impact of government action (i.e., total or partial lockdown), wearing the mask and respecting sanitary protocol such as the social distance. A typical example is a step-wise function; see Sect. 4.

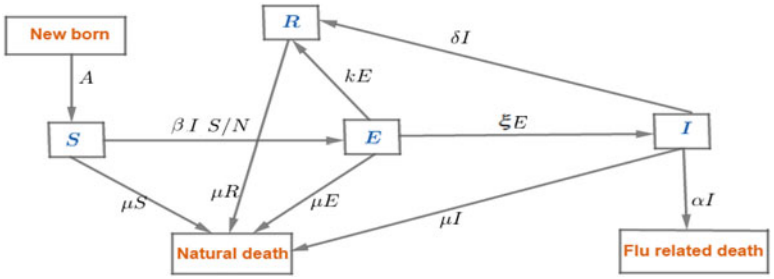


Fig. 1 The flow diagram of spatially homogeneous SIERD model

Table 1 Description of the parameters of the SEIRD system

Parameter	Description
A	Recruitment rate assumed $A = \mu N$
μ	Natural death rate for susceptible individuals
$\beta(t)$	Transmission rate function
ξ	Rate of progression from exposed to infectious
γ	Recovery rate of infectious individuals
α	Virus-induced average fatality rate

Assumptions 1–5 lead, by straightforward calculations, to the following model:

$$\begin{cases} \partial_t S = d_1 \nabla \cdot (\varphi_1(x) \nabla S) + A - \mu S - \beta(t) S \dfrac{I}{N}, \\ \partial_t E = d_2 \nabla \cdot (\varphi_2(x) \nabla E) + \beta(t) S \dfrac{I}{N} - (\mu + \xi) E, \\ \partial_t I = d_3 \nabla \cdot (\varphi_3(x) \nabla I) + \xi E - (\gamma + \mu + \alpha) I, \\ \partial_t R = d_4 \nabla \cdot (\varphi_4(x) \nabla R) + \gamma I - \mu R, \\ \partial_t D = \alpha I, \end{cases} \tag{1}$$

where $d_i,\ i = 1, \cdots, 4$ are the diffusion coefficients considered positive constants.

Mathematical model (1) is implemented with the following initial and boundary conditions:

$$\begin{cases} \dfrac{\partial S}{\partial \nu} = \dfrac{\partial E}{\partial \nu} = \dfrac{\partial I}{\partial \nu} = \dfrac{\partial R}{\partial \nu} = 0, & x \in \partial\Omega,\ t > 0, \\ S(0, x) = S_0(x),\ E(0, x) = E_0(x),\ I(0, x) = I_0(x), & x \in \Omega, \\ R(0, x) = R_0(x),\ D(0, x) = D_0(x), & x \in \Omega. \end{cases} \tag{2}$$

Assuming $\varphi_(x) = 1$ and $\beta(t) = \beta_0$ yields a spatially inhomogeneous SEIRD problem [30, 36]. However, if $d_i = 0$, system (1) reduces to a spatially homogeneous

SEIRD problem; see, e.g., [2, 17, 21, 22, 25, 33]. In this case, if $A = \mu = 0$ and $\xi = \infty$, system (1) is reduced to the classical SIR system.

Our chapter proposes a new model with additional features: (1) a transmission rate function rather than a constant and (2) the diffusion of individuals depends on a spatial distribution, which considers the preferred directions of propagation modeled by the functions $\varphi_i(x)$ for $i = 1, \ldots, 4$, which models the diffusion coefficient in the territory. Concretely, the model takes into account both transport and diffusion and, subsequently, the modeling of these terms takes into account the specific geography of the territory and, in particular, the transport network. Recent literature on this subject has been developed in [10–12]. We mention that the diffusion can be nonlinear. So, $\varphi_i = \varphi_i(x, h_i)$ for $i = 1, \ldots, 4$ where $h_1 = S$, $h_2 = E$, $h_3 = I$, and $h_4 = R$.

Considering a time-dependent transmission function can help to modeling the different strategies taken to defeat the virus, for instance, partial or total lockdown and the vaccination campaign. In general, the growth of the viral epidemic can be estimated by the basic reproduction ratio, denoted by R_0, which is the classical epidemiological measure associated with the reproductive power of the disease. For our system (1), it is given by the following function:

$$R_0(t) = \frac{\xi}{(\xi + \mu)(\gamma + \alpha + \mu)} \beta(t), \tag{3}$$

which provides a threshold for disease-free equilibrium point stability. Precisely, if $R_0(t) < 1$, the disease declines and disappears; while if $R_0(t) > 1$, an epidemic develops; see, e.g., [21, 40].

3 From Kinetic Theory Model to SEIRD Reaction–Diffusion System

Let us consider a multiscale approach to derive time-dependent SEIRD reaction–diffusion system (1) from a kinetic theory model by using the micro–macro decomposition method. First, we present its properties. Next, we rewrite it as a coupled system of microscopic part and macroscopic one and subsequently we derive macroscopic model by low order asymptotic expansions in terms of a small parameter.

3.1 Kinetic Theory Model

According to the parabolic–parabolic scaling limit [14], the kinetic theory model can be stated as follows for $i = 1, \ldots, 4$:

$$\begin{cases} \varepsilon \partial_t f_i + v \cdot \nabla_x f_i = \dfrac{1}{\varepsilon} \mathcal{T}_i[f_1, \cdots, f_{i-1}, f_{i+1}, \cdots, f_4](f_i) + \varepsilon\, G_i(f_1, \ldots, f_4), \\ \partial_t D = \alpha \displaystyle\int_V f_3\, dv, \\ f_i(0, x, v) = f_{i,0}(x, v), \qquad D(0, x) = D_0, \end{cases} \tag{4}$$

where $f_1(t, x, v), \cdots, f_4(t, x, v)$ are the distribution functions describing the statistical evolution of susceptible, exposed, infected, and recovered individuals, respectively. $t > 0$, $x \in \mathbf{R}^d$, $v \in V$ are, respectively, time, position, and velocity. The term $\mathcal{T}_i$ is the stochastic operator representing a random modification of direction of individuals and the operator G_i $(i = 1, \ldots, 4)$ describing the gain–loss balance of individuals. The mean free path ε measures the distance between individuals.

The micro–macro decomposition method is based on the assumptions in the following.

The turning operator $\mathcal{T}_i$ is decomposed as follows:

$$\mathcal{T}_i[f_1, \cdots, f_{i-1}, f_{i+1}, \cdots, f_4](f_i) = \mathcal{T}_i^1(f_i) + \varepsilon\, \mathcal{T}_i^2[f_1, \cdots, f_{i-1}, f_{i+1}, \cdots, f_4](f_i),$$

where $\mathcal{T}_i^1$ represents the dominant part of the turning kernel and is assumed to be independent of $f_1, \cdots, f_{i-1}, f_{i+1}, \cdots, f_4$. The operators $\mathcal{T}_i^j$ for $i = 1, \cdots, 4$ and $j = 1, 2$ are given by:

$$\mathcal{T}_i^j(f_i) = \int_V \big(T_i{}^j(v^*, v) f_i(t, x, v^*) - T_i{}^j(v, v^*) f_i(t, x, v)\big) dv^*, \tag{5}$$

where $T_i{}^j$ is the probability kernel for the new velocity $v \in V$ given that the previous velocity was v^*.

We assume that the operators $\mathcal{T}_i$ satisfy

$$\int_V \mathcal{T}_i\, dv = \int_V \mathcal{T}_i^1\, dv = \int_V \mathcal{T}_i^2\, dv = 0, \quad i = 1, \ldots, 4, \tag{6}$$

and that there exists a bounded velocity distribution $M_i(v) > 0$ independent of t and x such that

$$T_i^1(v, v^*) M_i(v^*) = T_i^1(v^*, v) M_i(v) \tag{7}$$

holds.

The flow produced by these equilibrium distributions vanishes and M_i are normalized, i.e.,

$$\int_V v\, M_i(v)dv = 0, \quad \int_V M_i(v)dv = 1, \quad i = 1, \ldots, 4. \tag{8}$$

Regarding the probability kernels, we assume that $T_i^1(v, v^*)$ is bounded, and there exist a constant $\sigma_i > 0$ ($i \in \{1, \ldots, 4\}$), such that

$$T_i^1(v, v^*) \geq \sigma_i M_i(v), \tag{9}$$

for all $(v, v^*) \in V \times V$, $x \in \Omega$ and $t > 0$.

Using the same arguments as in [3], the operator $\mathcal{T}_i^1$ has the following properties.

Lemma 1 *Assuming that hypothesis (6), (7), and (8) are satisfied. The following properties of the operators $\mathcal{T}_i^1$ for $i = 1, \ldots, 4$ hold true:*

(i) The operator $\mathcal{T}_i^1$ is self-adjoint in the space $\text{Ł}^2\left(V, \dfrac{dv}{M_i(v)}\right)$.

(ii) For $f \in \text{Ł}^2$, the equation $\mathcal{T}_i^1(g) = f$ has a unique solution $g \in \text{Ł}^2\left(V, \dfrac{dv}{M_i(v)}\right)$, satisfying

$$\int_V g(v)dv = 0 \quad \Longleftrightarrow \quad \int_V f(v)\, dv = 0.$$

(iii) The equation $\mathcal{T}_i^1(g) = v\, M_i(v)$, has a unique solution denoted by $\theta_i(v)$ for $i = 1, \ldots, 4$.

(iv) The kernel of $\mathcal{T}_i^1$ is $N(\mathcal{T}_i^1) = vect(M_i(v))$ for $i = 1, \ldots, 4$.

3.2 Micro–Macro Formulation

We rewrite the kinetic theory model (4) as a coupled system of microscopic part and macroscopic one. The main idea of the micro–macro method is to decompose the distribution function f_i for $i = 1, \ldots, 4$ as follows:

$$f_i(t, x, v) = M_i(v)u_i(t, x) + \varepsilon g_i(t, x, v),$$

where

$$u_i(t, x) = \langle f_i(t, x, v)\rangle := \int_V f_i(t, x, v)\, dv.$$

Then $\langle g_i\rangle = 0$ for $i = 1, \ldots, 4$. Inserting f_i in the kinetic model (4) and using the above-stated assumptions and properties of the interaction and the turning operators yields

$$\begin{cases} \partial_t(M_i(v)u_i) + \varepsilon\partial_t g_i + \dfrac{1}{\varepsilon} v M_i(v)\cdot\nabla u_i + v\cdot\nabla g_i = \dfrac{1}{\varepsilon}\mathcal{T}_i^1(g_i) \\ \\ \quad + \dfrac{1}{\varepsilon}\mathcal{T}_i^2[f_1,\cdots,f_{i-1},f_{i+1},\cdots,f_4](M_i u_i) \\ \\ \quad + \mathcal{T}_i^2[f_1,\cdots,f_{i-1},f_{i+1},\cdots,f_4](g_i) + G_i(f_1,\ldots,f_4), \\ \\ \partial_t D = \alpha u_3. \end{cases} \tag{10}$$

The micro–macro decomposition method is based on two steps. First, we use the projection technique to separate the macroscopic density $u_i(t,x)$ and microscopic quantity $g_i(t,x,v)$ for $i = 1,\ldots,4$. For that, we consider P_{M_i} the orthogonal projection onto $N(\mathcal{T}_i^1)$, for $i = 1,\ldots,4$. It follows

$$P_{M_i(v)}(h) = \langle h\rangle M_i(v), \quad \text{for any} \quad h \in \text{Ł}^2\left(V, \frac{dv}{M_i(v)}\right), \qquad i = 1,\ldots,4.$$

Inserting the operators $I - P_{M_i}$ into Eq. (10), using known properties for the projection P_{M_i} $i = 1,\cdots,4$, and integrating this equation with respect to the variable v yields the following micro–macro formulation:

$$\begin{cases} \partial_t g_i + \dfrac{1}{\varepsilon^2} v M_i(v)\cdot\nabla u_i + \dfrac{1}{\varepsilon}(I - P_{M_i})(v\cdot\nabla g_i) = \dfrac{1}{\varepsilon^2}\mathcal{T}_i^1(g_i) \\ \\ \quad + \dfrac{1}{\varepsilon}\mathcal{T}_i^2[f_1,\cdots,f_{i-1},f_{i+1},\cdots,f_4](M_i u_i) \\ \\ \quad + \mathcal{T}_i^2[f_1,\cdots,f_{i-1},f_{i+1},\cdots,f_4](g_i) + \dfrac{1}{\varepsilon}(I - P_{M_i})G_i(f_1,\ldots,f_4), \\ \\ \partial_t u_i + \langle v\cdot\nabla g_i\rangle = \langle G_i(f_1,\ldots,f_4)\rangle, \\ \\ \partial_t D = \alpha u_3. \end{cases} \tag{11}$$

The following proposition shows that micro–macro formulation (11) is equivalent to kinetic model (4).

Proposition 1

(i) Let $(f_1,\ldots,f_4)$ be a solution of kinetic theory model (4). Then $(u_1,\ldots,u_4,g_1,\ldots,g_4)$ is a solution of micro–macro formulation (11) associated with the following initial data for $i = 1,\ldots,4$:

$$u_i(t=0) = u_{i,0} = \langle f_{i,0}\rangle, \quad g_i(t=0) = g_{i,0} = \frac{1}{\varepsilon}(f_{i,0} - M_i u_{i,0}). \tag{12}$$

(ii) Conversely, if $(u_1, \ldots, u_4, g_1, \ldots, g_4)$ *is a solution of micro–macro formulation (11) associated with the following initial data* $(u_{1,0}, \ldots, u_{4,0}, g_{1,0}, \ldots, g_{4,0})$ *such that* $\langle g_{i,0} \rangle = 0$*, then* $(f_1, \ldots, f_4)$ *is a solution of the kinetic theory model (4) with initial data* $f_{i,0} = M_i u_{i,0} + \varepsilon g_{i,0}$ *and we have* $u_i = \langle f_i \rangle$ *and* $\langle g_i \rangle = 0$*, for* $i = 1, \ldots, 4$.

In order to develop asymptotic analysis of system (11), we assume that the interacting operators $\mathcal{T}_i^2$ and G_i satisfy the following asymptotic behavior in the limit

$$\mathcal{T}_i^2[M_1u_1 + \varepsilon g_1, \ldots, M_{i-1}u_{i-1} + \varepsilon g_{i-1}, M_{i+1}u_{i+1} + \varepsilon g_{i+1}, \ldots, M_4u_4 + \varepsilon g_4] : = \mathcal{T}_i^2[M_1u_1, \ldots, M_{i-1}u_{i-1}, M_{i+1}u_{i+1}, \ldots, M_4u_4] + O(\varepsilon), \tag{13}$$

and

$$G_i\Big(M_1(v)u_1 + \varepsilon g_1, \ldots, M_4(v)u_4 + \varepsilon g_4\Big) = G_i\Big(M_1(v)u_1, \ldots, M_4(v)u_4\Big) + O(\varepsilon), \tag{14}$$

for $i = 1, \ldots, 4$. The micro–macro formulation (11), which is equivalent to kinetic theory model (4), allows to obtain a general macroscopic model as ε goes to 0. Indeed, using (14) and (11), we obtain for $i = 1, \ldots, 4$

$$\mathcal{T}_i^1(g_i) = vM_i(v) \cdot \nabla u_i - \mathcal{T}_i^2[M_1u_1, \ldots, M_{i-1}u_{i-1}, M_{i+1}u_{i+1}, \ldots, M_4u_4](M_iu_i).$$

From Lemma 1, property (ii), the operator $\mathcal{T}_i$ is invertible. This implies

$$g_i = \big(\mathcal{T}_i^1\big)^{-1}\Big(vM_i \cdot \nabla u_i - \mathcal{T}_i^2[M_1u_1, \ldots, M_{i-1}u_{i-1}, M_{i+1}u_{i+1}, \ldots, M_4u_4] (M_iu_i)\Big). \tag{15}$$

Inserting (15) into the second equation in (11) yields the following macroscopic system:

$$\begin{cases} \partial_t u_i + \Big\langle v \cdot \nabla (\mathcal{T}_i^1)^{-1}\Big(vM_i(v) \cdot \nabla u_i \\ \qquad\qquad -\mathcal{T}_i^2[M_1u_1, \ldots, M_{i-1}u_{i-1}, M_{i+1}u_{i+1}, \ldots, M_4u_4](M_iu_i)\Big)\Big\rangle \\ \qquad\qquad = \Big\langle G_i(M_1(v)u_1, \ldots, M_4(v)u_4)\Big\rangle + O(\varepsilon), \\ \partial_t D = \alpha u_3. \end{cases} \tag{16}$$

Thanks to the following equalities:

$$\Big\langle v\cdot\nabla(\mathcal{T}_i^1)^{-1}\Big(vM_i(v)\cdot\nabla u_i\Big)\Big\rangle=\nabla\cdot\Big(\langle v\otimes\theta_i(v)\rangle\cdot\nabla u_i\Big),$$

and

$$\begin{aligned}&\Big\langle v\cdot\nabla(\mathcal{T}_i^1)^{-1}\Big(\mathcal{T}_i^2[M_1u_1,\dots,M_{i-1}u_{i-1},M_{i+1}u_{i+1},\dots,M_4u_4](M_iu_i)\Big)\Big\rangle\\&=\nabla\cdot\Big\langle\tfrac{\theta_i(v)}{M_i(v)}u_i\mathcal{T}_i^2[M_1u_1,\dots,M_{i-1}u_{i-1},M_{i+1}u_{i+1},\dots,M_4u_4](M_i)\Big\rangle,\end{aligned}$$

where $\theta_i(v)$ are given in Lemma 1 for $i=1,\dots,4$, we obtain the following general reaction–diffusion system:

$$\begin{cases}\partial_t u_i+\nabla\cdot\Big(\Gamma_i\big(u_1,\dots,u_{i-1},u_{i+1},\dots,u_4\big)u_i-d_i\cdot\nabla u_i\Big)\\\qquad\qquad=H_i(u_1,\dots,u_4)+O(\varepsilon),\\\partial_t D=\alpha u_3,\end{cases}\tag{17}$$

where d_i and the functions Γ_i and H_i are given by:

$$d_i=-\big\langle v\otimes\theta_i(v)\big\rangle,\tag{18}$$

$$\Gamma_i=-\Big\langle\frac{\theta_i(v)}{M_i(v)}u_i\mathcal{T}_i^2[M_1u_1,\dots,M_{i-1}u_{i-1},M_{i+1}u_{i+1},\dots,M_4u_4](M_i)\Big\rangle,\tag{19}$$

$$H_i(u_1,\dots,u_4)=\Big\langle G_i(M_1(v)u_1,\dots,M_4(v)u_4)\Big\rangle,\ \text{for } i=1,\dots,4.\tag{20}$$

To derive system (24), we consider in (4) specific choices of the appeared terms in the kinetic theory model (4). Namely,

$$u_1=S,\quad u_2=E,\quad u_3=I,\quad u_4=R.$$

The probability kernel T_i^1 is given by:

$$T_i^1=\frac{\sigma_i}{M_i(v)},\quad \text{for } i=1,\dots,4.$$

This implies

$$\mathcal{T}_i^1(g)=-\sigma_i\Big(g-M_i(v)\langle g\rangle\Big)=-\sigma_i\ g\quad\text{for } i=1,\dots,4.\tag{21}$$

Using (8), (21) and Lemma 1, then θ_i is given by:

$$\theta_i = -\frac{1}{\sigma_i} v M_i(v).$$

The other probability kernel T_i^2 is given by:

$$T_i^2 = \frac{\sigma_i\, d_i\, M_i\, v}{f_i}\big(1+\varphi_i(x)\big)\cdot\nabla\Big(\frac{f_i}{M_i}\Big), \quad \text{for } i=1,\ldots,4.$$

From (5) and Eq. (19), we obtain

$$\Gamma_i = \frac{d_i}{u_i}\big(1+\varphi_i(x)\big)\cdot\nabla u_i.$$

The modeling of the interaction operators G_i is given by:

$$\begin{cases} G_1(f_1,\ldots,f_4) = \dfrac{1}{|V|}\big(A-\mu f_1-\beta(t) f_1\dfrac{f_3}{n}\big),\\ G_2(f_1,\ldots,f_4) = \dfrac{1}{|V|}\big(\beta(t) f_1\dfrac{f_3}{n}-(\mu+\xi) f_2\big),\\ G_3(f_1,\ldots,f_4) = \dfrac{1}{|V|}\big(\xi f_2-(\gamma+\mu+\alpha) f_3\big),\\ G_4(f_1,\ldots,f_4) = \dfrac{1}{|V|}\big(\gamma f_3-\mu f_4\big), \end{cases} \tag{22}$$

where $n = 1-\sum_{i=1}^{4} f_i - D$. Then, using the definition of H_i in (20) to obtain from (22) the following equality:

$$\begin{cases} H_1(S,E,I,R) = A-\mu S-\beta(t) S\dfrac{I}{N},\\ H_2(S,E,I,R) = \beta(t) S I\dfrac{I}{N}-(\mu+\xi)E,\\ H_3(S,E,I,R) = \xi E-(\gamma+\mu+\alpha)I,\\ H_4(S,E,I,R) = \gamma I-\mu R. \end{cases} \tag{23}$$

Collecting the previous results, we obtain the time-dependent SEIRD reaction–diffusion system (1) of the order $O(\varepsilon)$

$$\begin{cases} \partial_t S = d_1 \nabla \cdot (\varphi_1(x) \nabla S) + A - \mu S - \beta(t) S \dfrac{I}{N} + O(\varepsilon), \\ \partial_t E = d_2 \nabla \cdot (\varphi_2(x) \nabla E) + \beta(t) S \dfrac{I}{N} - (\mu + \xi) E + O(\varepsilon), \\ \partial_t I = d_3 \nabla \cdot (\varphi_3(x) \nabla I) + \xi E - (\gamma + \mu + \alpha) I + O(\varepsilon), \\ \partial_t R = d_4 \nabla \cdot (\varphi_4(x) \nabla R) + \gamma I - \mu R + O(\varepsilon), \\ \partial_t D = \alpha I. \end{cases} \tag{24}$$

4 Numerical Method

We adopt an asymptotic preserving numerical scheme in one dimension based on the micro–macro method developed in Sect. 3. This method designs uniform stability with respect to the parameter ε, related to the mean distance between particles, as well as consistency with the reaction–diffusion limit. The discretization of micro–macro formulation (11) is carried out with respect to each independent variable (time, space, and velocity).

4.1 Semi-Implicit Time Discretization

We present a time discretization of micro–macro formulation (11). We denote by Δt a fixed time step, and by t_k a discrete time step such that $t_k = k\,\Delta t\ k \in N$. The approximation of $u_i(t, x)$ and $g_i(t, x, v)$ at the time step t_k are denoted, respectively, by $u_i^k \approx u_i(t_k, x)$ and $g_i^k \approx g_i(t_k, x, v)$.

In the first microscopic equations of (11), the only term that presents a stiffness in the collision part, for small ε, is $\dfrac{1}{\varepsilon}\mathcal{T}_i^1(g_i)$. Hence, we take an implicit scheme to ensure the stability for this term, while the other terms are still explicit. Consequently,

$$\begin{aligned} \frac{g_i^{k+1} - g_i^k}{\Delta t} + \frac{1}{\varepsilon^2} v M_i \cdot \nabla u_i^k + \frac{1}{\varepsilon}(I - P_{M_i})(v \cdot \nabla g_i^k) = \frac{1}{\varepsilon^2}\mathcal{T}_i^1(g_i^{k+1}) \\ + \frac{1}{\varepsilon^2}\mathcal{T}_i^2[M_1 u_1^k, \ldots, M_{i-1} u_{i-1}^k, M_{i+1} u_{i+1}^k, \ldots, M_4 u_4^k](M_i u_i^k) \\ + \frac{1}{\varepsilon}\mathcal{T}_i^2[M_1 u_1^k, \ldots, M_{i-1} u_{i-1}^k, M_{i+1} u_{i+1}^k, \ldots, M_4 u_4^k](g_i^k) \\ + \frac{1}{\varepsilon}(I - P_{M_i}) G_i(u_1^k, u_2^k, u_3^k, u_4^k). \end{aligned} \tag{25}$$

In the second macroscopic equations of (11), we take the function g at the time t_{k+1}, which gives

$$\frac{u_i^{k+1}-u_i^k}{\Delta t}+\langle v\cdot\nabla g_i^{k+1}\rangle=\Big\langle G_i(u_1^k,u_2^k,u_3^k,u_4^k)\Big\rangle. \tag{26}$$

Proposition 2 *The time discretization (25) and (26) is consistent with (16) in the limit.*

4.2 Fully Discrete Asymptotic Preserving Numerical Scheme in 1D

We construct a suitable space discretization of (25) and (26). The domain under consideration is $[-a,a]$. The velocity space $[-V,V]$ can be treated by using a standard discretization.

We define a staggered grid $x_j=j\,\Delta x$, $j=0,\ldots,N_x$ with $N_x=\dfrac{a}{\Delta x}$, and the cell center points $x_{j-\frac{1}{2}}=(j-\frac{1}{2})\Delta x$, $j=0,\ldots,N_x+1$. Let $u_{i,j}^k$ and $g_{i,j-\frac{1}{2}}^k$ be approximations of $u_i(t_k,x_j)$ and $g_i(t_k,x_{j-\frac{1}{2}},v)$, respectively. The microscopic equation (25) is discretized at points $x_{j+\frac{1}{2}}$, while the other macroscopic equations (26)) are discretized at point x_j. Then, we obtain

$$\frac{g_{i,j+\frac{1}{2}}^{k+1}-g_{i,j+\frac{1}{2}}^k}{\Delta t}+\frac{1}{\varepsilon^2}vM(v)\frac{u_{i,j+1}^k-u_{i,j}^k}{\Delta x}+\frac{1}{\varepsilon}(I-P_{M_i})\Big(v^+\frac{g_{i,j+\frac{1}{2}}^k-g_{i,j-\frac{1}{2}}^k}{\Delta x}$$

$$+v^-\frac{g_{i,j+\frac{3}{2}}^k-g_{i,j+\frac{1}{2}}^k}{\Delta x}\Big)=\frac{1}{\varepsilon^2}\mathcal{T}_i^1(g_{i,j+\frac{1}{2}}^{k+1})+\frac{1}{\varepsilon}(I-P_{M_i})G_i(u_{1,j+\frac{1}{2}}^k,\cdots,u_{4,j+\frac{1}{2}}^k)$$

$$+\frac{1}{\varepsilon^2}\mathcal{T}_i^2[M_1u_{1,j+\frac{1}{2}}^k,\ldots,M_{i-1}u_{i-1,j+\frac{1}{2}}^k,M_{i+1}u_{i+1,j+\frac{1}{2}}^k,\ldots,M_4u_{4,j+\frac{1}{2}}^k](M_iu_{i,j+\frac{1}{2}}^k)$$

$$+\frac{1}{\varepsilon}\mathcal{T}_i^2[M_1u_{1,j+\frac{1}{2}}^k,\ldots,M_{i-1}u_{i-1,j+\frac{1}{2}}^k,M_{i+1}u_{i+1,j+\frac{1}{2}}^k,\ldots,M_4u_{4,j+\frac{1}{2}}^k](g_{i,j+\frac{1}{2}}^k),$$

and

$$\frac{u_{i,j}^{k+1}-u_{i,j}^k}{\Delta t}+\Big\langle v\frac{g_{i,j+\frac{1}{2}}^{k+1}-g_{i,j-\frac{1}{2}}^{k+1}}{\Delta x}\Big\rangle=\langle G_i(u_{1,j}^k,\cdots,u_{4,j}^k)\rangle,$$

where $u_{i,j+\frac{1}{2}}=\frac{u_{i,j+1}+u_{i,j}}{\Delta x}$.

Proposition 3 *The time and space approximation in 4.2 of kinetic equation (4) in the limit ε goes to zero satisfy the following discretization:*

$$\frac{u_i^{k+1}-u_i^k}{\Delta t}+\frac{1}{\Delta x}\Big\langle v\cdot\Big[(\mathcal{T}_i^1)^{-1}\Big(vM(v)\frac{u_{i,j+1}^k-u_{i,j}^k}{\Delta x}+vM(v)\frac{u_{i,j}^k-u_{i,j-1}^k}{\Delta x}$$

$$-\mathcal{T}_i^2[M_1u^k_{1,j+\frac{1}{2}},\dots,M_{i-1}u^k_{i-1,j+\frac{1}{2}},M_{i+1}u^k_{i+1,j+\frac{1}{2}},\dots,M_4u^k_{4,j+\frac{1}{2}}](M_iu^k_{i,j+\frac{1}{2}})$$

$$-\mathcal{T}_i^2[M_1u^k_{1,j-\frac{1}{2}},\dots,M_{i-1}u^k_{i-1,j-\frac{1}{2}},M_{i+1}u^k_{i+1,j-\frac{1}{2}},\dots,M_4u^k_{4,j-\frac{1}{2}}](M_iu^k_{i,j-\frac{1}{2}})\Big)\Big]\Big\rangle$$

$$=\langle G_i(u^k_{1,j+\frac{1}{2}},\cdots,u^k_{4,j+\frac{1}{2}})\rangle,$$

which is consistent with the first equation of (15).

4.3 Boundary Conditions

For the numerical solution of the kinetic equation (1), usually the inflow boundary conditions are prescribed as follows:

$$f_i(t,x_{\min},v)=f_{i,l}(v),\quad v>0,\qquad f_i(t,x_{\max},v)=f_{i,r}(v),\quad v<0,$$
$$\text{for } i=1,\dots,4.$$

Thus, the inflow boundary conditions can be rewritten in the micro–macro formulation (11) as follows:

$$u_i(t,x_0)M_i(v)+\frac{\varepsilon}{2}(g_i(t,x_{\frac{1}{2}},v)+g_i(t,x_{-\frac{1}{2}},v))=f_{i,l}(v),\quad v<0,$$

$$u_i(t,x_{N_x})M_i(v)+\frac{\varepsilon}{2}(g_i(t,x_{N_x+\frac{1}{2}},v)+g_i(t,x_{N_x-\frac{1}{2}},v))=f_{i,r}(v),\quad v>0.$$

We consider the following artificial Neumann boundary conditions for the other velocities:

$$g_i(t,x_{\frac{1}{2}},v)=g_i(t,x_{-\frac{1}{2}},v),\quad v<0,$$

$$g_i(t,x_{N_x+\frac{1}{2}},v)=g_i(t,x_{N_x-\frac{1}{2}},v),\quad v>0.$$

Furthermore, the ghost points can be computed as follows:

$$g_{i,j-\frac{1}{2}}^{k+1} = \begin{cases} \dfrac{2}{\varepsilon}\Big(f_{i,l}(v) - u_{i,0}^{k+1} M_i(v)\Big) - g_{i,\frac{1}{2}}^{k+1}, & v > 0, \\ g_{i,\frac{1}{2}}^{k+1}, & v < 0, \end{cases} \tag{27}$$

and

$$g_{i,N_x+\frac{1}{2}}^{k+1} = \begin{cases} \dfrac{2}{\varepsilon}\Big(f_{i,r}(v) - u_{i,N_x}^{k+1} M_i(v)\Big) - g_{i,N_x-\frac{1}{2}}^{k+1}, & v < 0, \\ g_{i,N_x-\frac{1}{2}}^{k+1}, & v > 0. \end{cases} \tag{28}$$

Finally, from (4.2) we have

$$\begin{cases} \Big(1 + \dfrac{2\Delta t}{\varepsilon \Delta x}\langle v^+ M_i(v)\rangle\Big) u_{i,0}^{k+1} = u_{i,0}^k - \dfrac{\Delta t}{\Delta x}\Big\langle (v + v^+ - v^-) g_{i,\frac{1}{2}}^{k+1} \\ \qquad\qquad + - \dfrac{2v_l^+}{\varepsilon} f_l(v)\Big\rangle \Delta t\, G_i(u_{1,0}^k, \cdots, u_{4,0}^k), \\ \Big(1 - \dfrac{2\Delta t}{\varepsilon \Delta x}\langle v^- M_i(v)\rangle\Big) u_{i,N_x}^{k+1} = u_{i,N_x}^k - \dfrac{\Delta t}{\Delta x}\Big\langle \dfrac{2v^-}{\varepsilon} f_r(v) \\ \qquad\qquad -(v - v^+ + v^-) g_{i,N_x-\frac{1}{2}}^{k+1}\Big\rangle + \Delta t\, G_i(u_{1,N_x}^k, \cdots, u_{4,N_x}^k). \end{cases} \tag{29}$$

5 Numerical Results

In this section, we provide some numerical simulations obtained both with the micro–macro formulation presented in Sect. 4 and from the macroscopic scheme. First, we show the asymptotic preservation numerical scheme property. Second, we demonstrate the effect of the diffusion terms on the evolution of individuals and its sensitivity for the different choices of the reproduction ratio R_0. Finally, we show how the transmission function $\beta(t)$ influences the evolution of the pandemic.

We consider that the velocity space is the interval $V = [-1.1]$ with the number of grids $N_v = 164$, which can provide sufficient precision for numerical simulations [15]. The step time is $t = 10^{-3}$ and the space domain is the interval $\Omega = [-2, 2]$ with the number of cells $N_x = 200$. We take the following set of parameters as an example to analyze the results by varying some of them: $\mu = 1/83,\ \alpha = 0.006,\ \xi = 1/4, and\ \gamma = 1/8$. Three cases of the diffusion coefficients are considered: (a) without diffusion ($d_i = 0$), (b) with diffusion, namely $d_1 = 0.05,\ d_2 = 0.025,\ d_3 = 0.001$, and $d_4 = 0$, the same as Reference

[36] where the functions $\varphi_i(x) = 1$, and (c) same as the case (ii) with diffusion coefficient depending on x, namely, $\varphi_i(x) =\mid x \mid$. Finally, we take the following initial conditions:

$$(i)\begin{cases} S_0 = 2.6\left(\exp(-(\frac{x-0.5}{0.12})^2) + \exp(-(\frac{x+0.5}{0.12})^2)\right)/(0.9\,\pi), \\ I_0 = 0.04\,\exp(-2\,x^2), \\ E_0 = R_0 = 0, \end{cases}$$

and

$$(ii)\begin{cases} S_0 = 0.96\,\exp(-10(\frac{x}{1.4})^2), \\ I_0 = 0.04\,\exp(-2\,x^2), \\ E_0 = R_0 = 0, \end{cases}$$

5.1 Test 1: Asymptotic Preserving Numerical Scheme Property

In this test we aim to validate the asymptotic preserving numerical scheme property. We consider the initial conditions (i), the diffusion case (b), and the reproduction ratio is $R_0 = 2$. In Fig. 2, we show the numerical results of susceptible, exposed, and infected individuals obtained with micro–macro scheme presented in Sect. 4 and with macroscopic numerical scheme at successive instants $t = 0.5,\ 1,\ 5,\ 10$. We observe that the obtained numerical results have almost the same profiles in the limit when the parameter $\varepsilon = 2 \times 10^{-k}$, with $k = 0,\ 1,\ 2,\ 3,\ 4,\ 6$, goes to zero. This confirms that the asymptotic preserving numerical scheme is uniformly stable along the transition from kinetic regime to macroscopic regime, which illustrates the result in Proposition 3.

5.2 Test 2: Diffusion Effect

In this test we demonstrate the effect of diffusion over the interacting individuals. We consider the initial conditions (ii) and the reproduction ratio is $R_0 = 2$. Figure 3 shows the numerical results of susceptible, exposed, and infected individuals obtained with asymptotic preserving scheme where $\varepsilon = 10^{-6}$ in three cases: without diffusion, case a) illustrated in sub-figures (a)–(c), with diffusion case (b) illustrated in sub-figures (d)–(f), and with diffusion case (c) illustrated in sub-figures (g)–(i). We observe that the individuals are all centered around the axis $x = 0$ in the absence of diffusion. Whereas, when we consider diffusion with $\varphi_i = 1$, it is clear that individuals are more spreading within the domain. In the last case, where $\varphi_i(x) =\mid x \mid$, we notice that the spreading is limited around the axis $x = 0$.

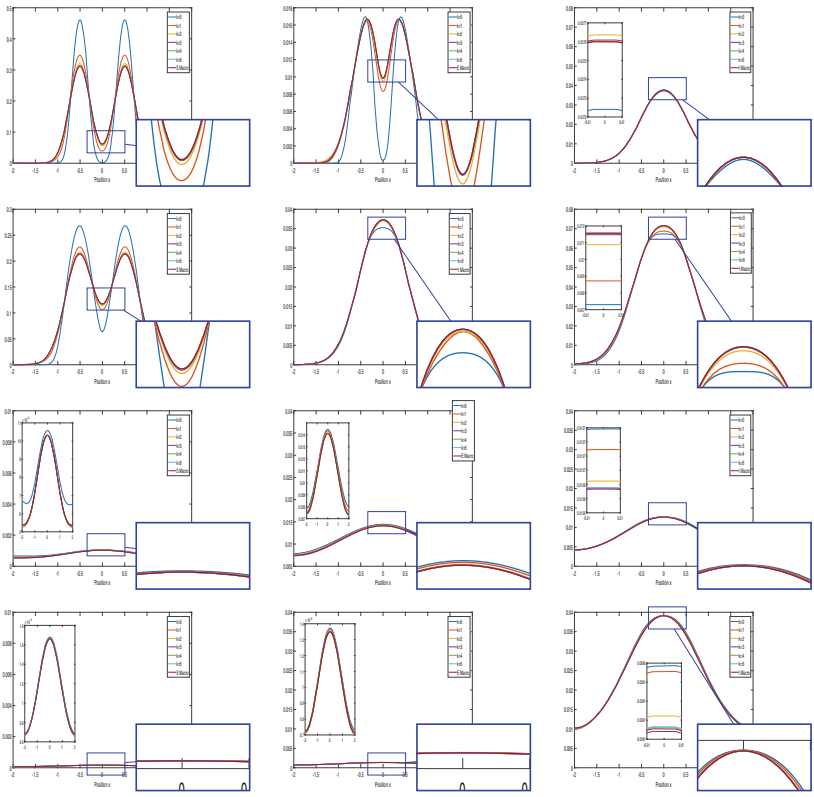


Fig. 2 Dynamics of the densities f_1 (first column), f_2 (second column), and f_3 (third column) obtained with the asymptotic preserving numerical scheme for $\varepsilon = 2 \times 10^{-k}$, $k = 0,\ 1,\ 2,\ 3,\ 4,\ 6$ and with the macroscopic numerical scheme using initial conditions (i) at successive time $t = 0.5,\ 1,\ 5,\ 10$

5.3 Test 3: Role of the Transmission Function

Here we show how the transmission rate function influences the evolution of the pandemic. First, we consider constant values of $\beta = 0.03,\ 0.075,\ 0.1799,\ 0.7497,\ 1.12,\ 2.2491$ (the corresponding reproduction ratio is $R_0 = 0.2,\ 0.5,\ 0.8,\ 1.2,\ 5,\ 15$, respectively). Figure 4 shows the variation over time of susceptible, exposed, and infected individuals with diffusion case (b) at $x = 0$ performed with the initial condition (i). We observe that for small values of the transmission rate, the proportion of infected individuals is low. In addition, steady state results in a relatively small proportion of the population in the recovered individuals, while the bulk of the population remains in the susceptible individuals (who have not caught the disease). However, for relatively high and moderate values of β, a large proportion of individuals are found at equilibrium in recovered individuals. This

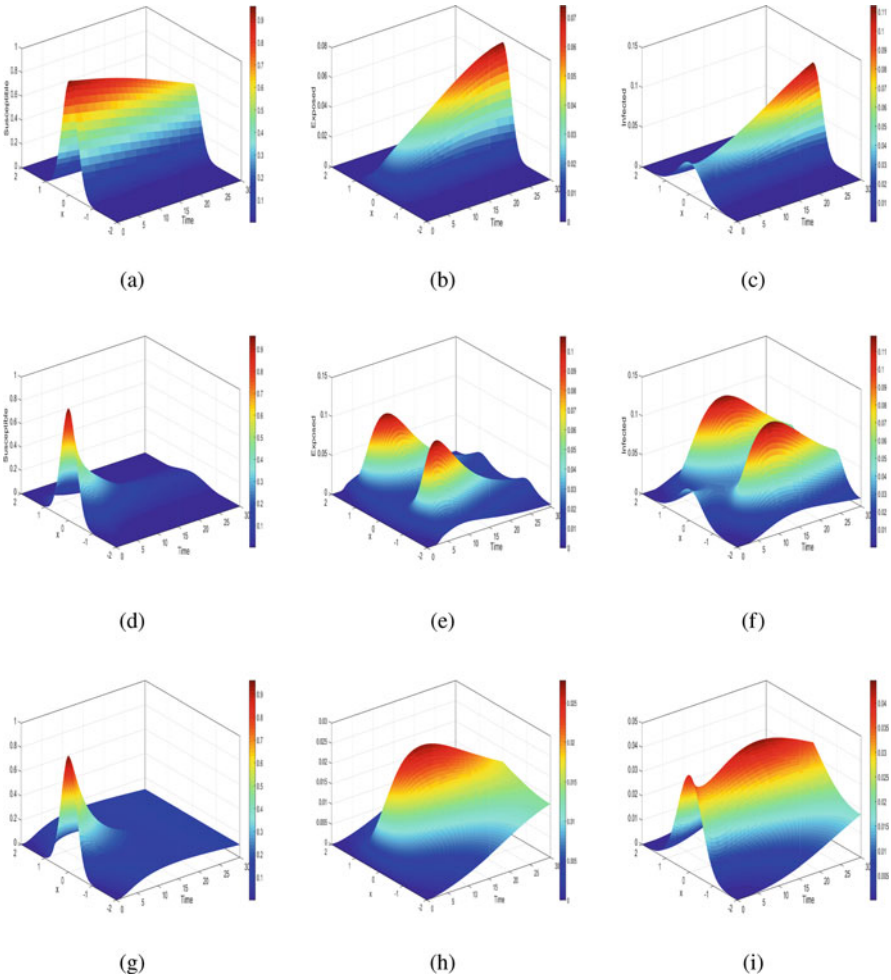


Fig. 3 Evolution of densities f_1, f_2, f_3 obtained with the asymptotic preserving numerical scheme for $\varepsilon = 10^{-6}$ and initial condition (ii): without diffusion case (a) (sub-figures (**a**), (**b**), (**c**)), with diffusion case (b) (sub-figures (**d**), (**e**), (**f**)), and with diffusion case (c) (sub-figures (**g**), (**h**), (**i**)). The reproduction ratio is $R_0 = 2$

means that most of the population caught the disease and got infected, and then recovered. Note that in this case only a relatively small proportion of the population remains in the susceptible individuals. Finally, we notice that infected and exposed individuals disappear after a reasonable period of time, while susceptible and recovered individuals reach a non-zero constant at steady state.

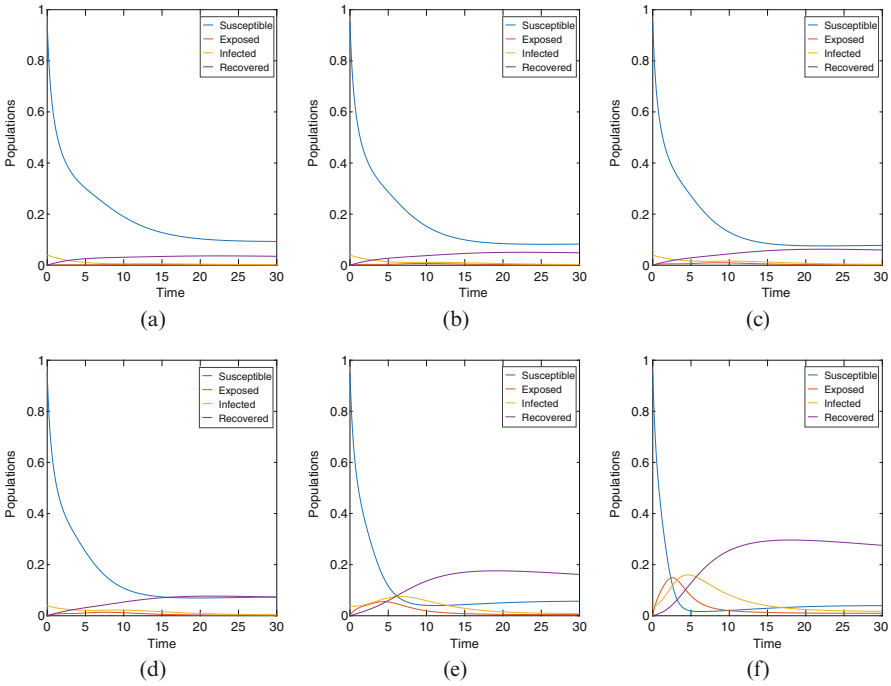


Fig. 4 Time variation of the numerical results obtained with the asymptotic preserving scheme using initial condition (ii), $\varepsilon = 10^{-6}$, and diffusion case (a) at $x = 0$. The transmission rate values are $\beta = 0.03,\ 0.075,\ 1.12,\ 0.1799,\ 0.7497,\ 2.2491$

Now, let us consider a time-dependent transmission rate $\beta(t)$ giving by the following step-wise function:

$$\beta(t) = 0.121\!\!1_{[0,T/3]}(t) + 2.24911\!\!1_{]T/3,2T/3]}(t) + 0.0151\!\!1_{]2T/3,T]}(t), \tag{30}$$

where $T = 100$.

Figure 5 shows time variation of infected and died individuals obtained with the asymptotic preserving numerical scheme with $\varepsilon = 10^{-6}$, diffusion case (b), and initial condition (ii) at $x = 0$. It is clear that the numbers of infected and died individuals increase from time $t = 10$. In fact, this occurs because of the value of β, which corresponds to $R_0(t) = 15 > 1$. As time progresses, we notice that the numbers of infected and died individuals decrease at time $T > 20$. This happens thanks to the choice of the transmission rate function, where a small value of $R_0(t) = 0.2 < 1$ is considered.

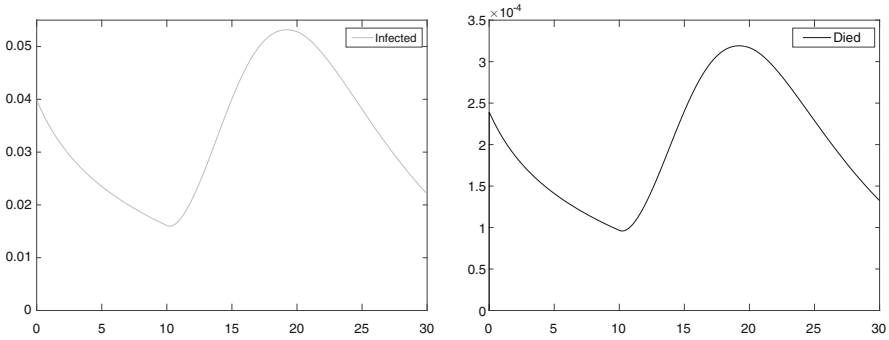


Fig. 5 Time variation of infected and died individuals obtained with asymptotic preserving numerical scheme using initial condition (ii) and diffusion case (b), at $x = 0$. The transmission rate function $\beta(t)$ is given by Eq. (30) and $\varepsilon = 10^{-6}$

6 Conclusion and Perspectives

In this chapter, a time-independent Susceptible-Exposed-Infectious-Recovered-Died reaction–diffusion system for the virus pandemic has been proposed. This system is derived from kinetic theory model using the multiscale approach.

Various promising numerical simulations have been provided in three tests. First, the asymptotic preserving numerical scheme has been shown to be uniformly stable along the transition from the kinetic regime to the macroscopic regime. Second, how the diffusion terms affect the spreading of the pandemic has been performed with three cases. Finally, the role of the transmission rate function has been demonstrated.

We think that this chapter opens interesting perspectives such as extension of the proposed macroscopic model by adding cross-diffusion terms, and looking for applications in other fields, e.g., the mechanism of normal and infected cells [7, 38].

References

1. https://www.worldometers.info/coronavirus. Accessed 17 Jan 2021
2. F. Al-Showaikh, E. Twizell, One-dimensional measles dynamics. Appl. Math. Comput. **152**, 169–194 (2004)
3. A. Atlas, M. Bendahmane, F. Karami, D. Meskine, M. Zagour, Kinetic-fluid derivation and mathematical analysis of nonlocal cross-diffusion–fluid system. Appl. Math. Model. **82**, 379–408 (2020)
4. N.T.J. Bailey, *The Mathematical Theory of Infectious Diseases and its Applications*, 2nd edn. (Hafner Press, New York, 1975)
5. N. Bellomo, A. Bellouquid, J. Nieto, J. Soler, On the asymptotic theory from microscopic to macroscopic tissue models: an overview with perspectives. Math. Models Methods Appl. Sci. **22**, 1130001 (2012)

6. N. Bellomo, A. Bellouquid, N. Chouhad, From a multiscale derivation of nonlinear cross-diffusion models to Keller-Segel models in a Navier-Stokes fluid. Math. Models Methods Appl. Sci. **26**, 2041–2069 (2016)
7. N. Bellomo, K.J. Painter, Y. Tao, M. Winkler, Occurrence vs. absence of taxis-driven instabilities in a May–Nowak model for virus infection. SIAM J. Appl. Math. **79**(5), 1990–2010 (2019)
8. N. Bellomo, R. Bingham, M.A. Chaplain, G. Dosi, G. Forni, D.A. Knopoff, J. Lowengrub, R. Twarock, M.E. Virgillito, A multi-scale model of virus pandemic: heterogeneous interactive entities in a globally connected world. Math. Models Methods Appl. Sci. **30**(8), 1591–1651 (2020)
9. M. Bendahmane, F. Karami, M. Zagour, Kinetic-fluid derivation and mathematical analysis of the cross-diffusion–brinkman system. Math. Methods Appl. Sci. **41**(16), 6288–6311 (2018)
10. H. Berestycki, J.M. Roquejoffre, L. Rossi, Propagation of epidemics along lines with fast diffusion. Bull. Math. Biol. **83**(1), 2 (2021)
11. G. Bertaglia, L. Pareschi, Hyperbolic models for the spread of epidemics on networks: kinetic description and numerical methods. ESAIM: M2AN **55**(2), 381–407 (2021)
12. W. Boscheri, G. Dimarco, L. Pareschi, Modeling and simulating the spatial spread of an epidemic through multiscale kinetic transport equations. Math. Models Methods Appl. Sci. **31**(6), 1059–1097 (2021)
13. F. Brauer, Mathematical epidemiology: past, present, and future. Infect. Dis. Model. **2**(2), 113–127 (2017)
14. D. Burini, N. Chouhad, A multiscale view of nonlinear diffusion in biology: from cells to tissues. Math. Models Methods Appl. Sci. **29**(4), 791–823 (2019)
15. J.A. Carrillo, B. Yan, An asymptotic preserving scheme for the diffusive limit of kinetic systems for chemotaxis. Multiscale Model. Simul. **11**(1), 336–361 (2013)
16. G. Chowell, Fitting dynamic models to epidemic outbreaks with quantified uncertainty: a primer for parameter uncertainty, identifiability, and forecasts. Infect. Dis. Model. **2**(3), 379–398 (2017)
17. O. Diekmann, H. Heesterbeek, T. Britton, *Mathematical Tools for Understanding Infectious Disease Dynamics.* Princeton Series in Theoretical and Computational Biology (Princeton University Press, Princeton, 2013)
18. E. Estrada, COVID-19 and SARS-CoV-2. Modeling the present, looking at the future. Phys. Rep. **869**, 1–51 (2020)
19. G.J. Fox, J.M. Traue, E. McBryde, Modelling the impact of COVID?19 on intensive care services in New South Wales. Med. J. Aust. **212**, 468–469 (2020)
20. M. Gatto, E. Bertuzzo, L. Mari, S. Miccoli, L. Carraro, R. Casagrandi, A. Rinaldo, Spread and dynamics of the COVID-19 epidemic in Italy: effects of emergency containment measures. Proc. Natl. Acad. Sci. USA **117**(19), 10484–10491 (2020)
21. H.W. Hethcote, The mathematics of infectious diseases. SIAM Rev. **42**(4), 599–653 (2000)
22. C. Hou, J. Chen, Y. Zhou, L. Hua, J. Yuan, S. He, Y. Guo, S. Zhang, Q. Jia, C. Zhao, et al., The effectiveness of quarantine of Wuhan city against the Corona Virus Disease 2019 (COVID-19): a well-mixed SEIR model analysis. J. Med. Virol. **92**, 841–848 (2020)
23. S. Jin, Efficient asymptotic-preserving (AP) schemes for some multiscale kinetic equations. SIAM J. Sci. Comput. **21**(2), 441–454 (1999)
24. M.J. Keeling, K.T. Eames, Networks and epidemic models. J. R. Soc. Interface **2**(4), 295–307 (2005)
25. M.J. Keeling, P. Rohani, *Modeling Infectious Diseases in Humans and Animals* (Princeton University Press, Princeton, 2011)
26. A. Klar, Asymptotic-induced domain decomposition methods for kinetic and drift diffusion semiconductor equations. SIAM J. Sci. Comput. **19**, 2032–2050 (1998)
27. W. Kermack, A. McKendrick, A contribution to the mathematical theory of epidemics. Proc. R. Soc. Lond. A **115**, 700–721 (1927)

28. Q. Li, B. Tang, N.L. Bragazzi, Y. Xiao, J. Wu, Modeling the impact of mass influenza vaccination and public health interventions on COVID-19 epidemics with limited detection capability. Math. Biosci. **325**, 108378 (2020)
29. Q. Lin, S. Zhao, D. Gao,Y. Lou, S. Yang, S.S. Musa, M.H. Wang, Y. Cai, W. Wang, L. Yang, D. He, A conceptual model for the outbreak of Coronavirus disease 2019 (COVID-19) in Wuhan, China with individual reaction and governmental action. Int. J. Infect. Dis. **93**, 211–216 (2020)
30. E. Massad, M.N. Burattini, F.A.B. Coutinho, L.F. Lopez, The 1918 influenza a epidemic in the city of Sao Paulo, Brazil. Med. Hypotheses **68**(2), 442–445 (2007)
31. L. Pang, S. Liu, X. Zhang, T. Tian, Z. Zhao, Transmission dynamics and control strategies of covid-19 in Wuhan, China. J. Biol. Syst. **28**(3), 543–560 (2020)
32. B. Perthame, *Transport Equations in Biology* (Blackwell Science Ltd, Hoboken, 1998)
33. C.W. Potter, K.G. Nicholson, R.G. Webster, A.J. Hay, *Textbook of Influenza*. Birkhäuser, Series "Frontiers in biology" (Birkhäuser, Basel, 2007)
34. F.J. Richards, A flexible growth function for empirical use. J. Exp. Bot. **10**(2), 290–301 (1959)
35. G. Röst, SEIR epidemiological model with varying infectivity and infinite delay. Math. Biosci. Eng. **5**(2), 389–402 (2008)
36. M. Samsuzzoha, M. Singh, D. Lucy, Numerical study of an influenza epidemic model with diffusion. J. Appl. Math. Comput. **217**, 3461–3479 (2010)
37. N. Shao, M. Zhong, Y. Yan, H. Pan, J. Cheng, W. Chen, Dynamic models for coronavirus disease 2019 and data analysis. Math. Methods Appl. Sci. **43**, 4943–4949 (2020)
38. O. Stancevic, C.N. Angstmann, J.M. Murray, B.I. Henry, Turing patterns from dynamics of early HIV infection. Bull. Math. Biol. **75**, 774–795 (2013)
39. G.Q. Sun, S.F. Wang, M.T. Li, L. Li, J. Zhang, W. Zhang, Z. Jin, G.L. Feng, Transmission dynamics of COVID-19 in Wuhan, China: effects of lockdown and medical resources. Nonlinear Dyn. **101**, 1981–1993 (2020)
40. P. Van den Driessche, J. Watmough, Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Math. Biosci. **180**(1–2), 29–48 (2002)
41. P.F. Verhulst, Notice sur la loi que la population suit dans son accroissement. Corr. Math. Phys. **10**, 113 (1838)
42. J.O. Wertheim, A.J. Leigh Brown, N.L. Hepler, S.R. Mehta, D.D. Richman, D.M. Smith, S.L. Kosakovsky Pond, The global transmission network of HIV-1. J. Infect. Dis. **209**(2), 304–313 (2014)