



**HAL**  
open science

# Challenges for mathematical epidemiological modelling

Pascal Crepey, Harold Noël, Samuel Alizon

► **To cite this version:**

Pascal Crepey, Harold Noël, Samuel Alizon. Challenges for mathematical epidemiological modelling. *Anaesthesia Critical Care & Pain Medicine*, 2022, 41 (2), pp.101053. 10.1016/j.accpm.2022.101053 . hal-03607991

**HAL Id: hal-03607991**

**<https://ehesp.hal.science/hal-03607991>**

Submitted on 12 Apr 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

## Challenges for mathematical epidemiological modelling

Pascal CRÉPEY<sup>1</sup>, Harold NOËL<sup>2</sup>, Samuel ALIZON<sup>3,4\*</sup>

<sup>1</sup> ARENES – UMR 6051, Univ Rennes, EHESP, CNRS, Rennes, France

<sup>2</sup> Direction des maladies infectieuses, Santé Publique France, Saint-Maurice, France

<sup>3</sup> MIVEGEC, CNRS, IRD, Université de Montpellier, Montpellier, France

<sup>4</sup> Centre for Interdisciplinary Research in Biology (CIRB), College de France, CNRS, INSERM, Université PSL, Paris, France

### **\*Corresponding author**

**Email:** samuel.alizon@cnrs.fr

**Keywords:** nowcasting, forecasting, mathematics, statistics, individual-based models, data integration

### **Conflicts of interest:**

PC has received consulting fees from Sanofi Pasteur for projects unrelated to COVID-19.

### **Acknowledgments:**

SA thanks the CNRS, the IRD, and the Université of Montpellier for funding.

A century-old field

The 16<sup>th</sup> of March 2020 will probably go down in history as an important date for epidemiological modelling. On that day, the Imperial College COVID-19 response team

published a report investigating several scenarios exploring the impact of an epidemic wave in the UK and the USA [1]. Few modelling studies ever had this impact since it represented the main scientific basis to enforce national lockdowns in many countries. Such a rapid deployment of informative scenarios results from decades of research in mathematical epidemiology. The now classical *SIR* model, which follows the dynamics of individuals alternating between three different states (Susceptible, Infected, and Recovered), dates from a model by Kermack and McKendrick published in 1927 [2], which followed earlier work by Ross and Hudson [3]. Over a century, mathematical epidemiology blossomed, integrating new mathematical techniques (*e.g.*, partial differential equations to capture the infection age structure already present in Kermack and McKendrick's model, or Bayesian inference to improve parameter estimation) as well as detailed biological life cycles (*i.e.*, adding more compartments to the *SIR* model with, for instance, explicit vector-borne transmission, latent infection stages, or vaccination) and health interventions.

Mathematical models use data as an input but their output is typically more "hypothetical" than statistical models and referred to here as "scenarios". Their goal is to improve our understanding of a system by exploring the values of variables of interest for a given set of assumptions. Given the infinite number of possibilities, the choice of the hypotheses to test is essential. Epidemiological modelling is truly a multidisciplinary endeavour because, beyond mathematics, statistics, and computer science, it requires knowledge in biology, immunology, microbiology, and public health. Other fields, such as demographics, sociology, or economics, have also become increasingly important to address specific questions. Integrating this diversity is a challenge because each field comes with its own concepts and methods.

Although some epidemiological models are a century old, the field is extremely lively, as illustrated by a recent special issue on COVID-19 [4]. One might add “unfortunately” because its scientific advances correlate with major epidemics. For instance, a method routinely used to estimate the temporal reproduction number of the epidemic, *i.e.*, the average number of secondary infections caused by an infected person over the course of his/her infection was implemented to analyse field data from the Ebola epidemic in West Africa in 2013-2016 [4]. At the time, epidemiologists could build on modelling improvements made during the 2009 A/H1N1 pandemic. This ongoing progress comes from the fact that many features of an epidemic, especially the intensity of the measures required to control it, are difficult to forecast [6].

#### A deluge of data

Epidemiologists in the previous century struggled to extract signals from limited data but, nowadays, modellers face a deluge of data. The shift is patent in terms of quantity, accessibility (often in near-real time), and heterogeneity. Some data can readily be incorporated in most models (*e.g.*, incidence data), although the rapid availability can amplify some biases, such as reporting delays. Others, such as mobility data, can enrich existing spatial models but are more challenging to analyse and incorporate, and raise acute ethical issues [7]. Finally, some data like virus sequences require completely different frameworks, for example phylodynamics [8].

A challenge for modellers is known as “data integration” [9]. It hypothesises that combining different sources of data can reveal information that would be inaccessible using a single data source. For example, phylodynamics and genomics can achieve inferences of temporal

reproduction numbers that are sometimes comparable to those obtained using incidence data [10]. However, to convince public health agencies, models should demonstrate that using both incidence and virus genomic data can yield even more insights. Also, each data source brings its own biases and uncertainties, and it is still unclear how these interact. Therefore, data integration requires a high degree of multidisciplinary to avoid known pitfalls in each field.

### Integrating between scales

Diseases take their roots in molecular interactions and it is tempting to incorporate detailed within-patient processes into transmission models [11]. However, integrating multiple scales can lead to irrelevant model parameters and having an infected cell decay rate in a model describing nationwide hospital dynamics makes little sense. One aspect that can readily be included in population models is the within-host kinetics [12]. This allows models to account for variations of infectivity and, for some systems, of virulence [13].

The geographical scale raises similar issues. Infections spread at a local level but public health policies occur at a regional or national level. Early in an epidemic, this problem can be neglected because transmission chains are independent [14]. However, as the infection spreads or as vaccination coverage increases, local heterogeneities appear. This explains why in absence of intervention the magnitude of an epidemic peak can differ from that predicted using well-mixed models, where every individual can be in contact with one another, as shown by the EPIDEMAP framework [15] illustrated in **Figure 1**. This also impacts the optimal level of public health interventions. For instance, modelling suggests

that local (county-level) interventions can be as efficient as countrywide interventions in preventing SARS-CoV-2 infections while causing half the impact in terms of closing days [16]. The integration of multiple scales also matters for multi-host dynamics. The existence of animal or environmental reservoirs calls for a detailed understanding of the ecology of the parasite to capture stochastic spill overs [17] or specific dynamics such as dilution effects [18].

Finally, the temporal dimension overlaps with these multiscale issues. This is obvious for the within-host/between-host interface, but it is also true for the evolutionary scale. Over the last decade, models building on quantitative genetics showed how pathogen evolution could shape epidemiological dynamics [19]. After the evolution of SARS-CoV-2 “variants of concern”, which led to national lockdowns, evolutionary epidemiology stands as a key challenge for epidemiological models.

The increasing heterogeneity raises acute challenges for medium- or long-term models, for instance, to anticipate the impact of “long-COVIDs” or design vaccination campaigns. The increase in computational power allows us to simulate national dynamics with a high degree of resolutions [15], but achieving similar integration between scales using flexible mathematical models remains challenging.

#### Models for public health intervention

The COVID-19 pandemic called for control measures with unprecedented social and economic consequences, spawning new expectations about how mathematical models should inform policy decisions or interventions. As the pandemic continued, decision-makers were confronted with the need to constantly update measures such as social distancing, schools screening, or vaccine deployment. Statistical models could rank the

efficiency of interventions [20] but a more accurate forecasting of epidemic trajectories became essential to determine the timeliness, scope, duration, or type of interventions to implement. This led to the development of “nowcasting” approaches – *i.e.*, forecasting on a short-term scale period of a few days or weeks [4]. In theory, nowcasting complements the epidemiological surveillance for early warning signals and can help implement reactive custom-tailored interventions that match epidemic dynamics. However, it does raise important pitfalls. First, if there are many unknowns, the number of scenarios to explore is large, which can generate misunderstandings with public health or political officials who may select some scenarios or get lost among all the possibilities. Presenting many scenarios as equally likely may also erode public trust and willingness to base public health decisions on models. A promising avenue of research is to compute, for each scenario, a likelihood measure that allows updating the ranking between scenarios when new data becomes available. Finally, having multiple teams tackling the same question will always be informative and an easy way to encourage this is through forecasting contests such as <https://covid19forecasthub.eu/>, an open challenge, where modellers compare and aggregate models at the European level.

### Communicating models

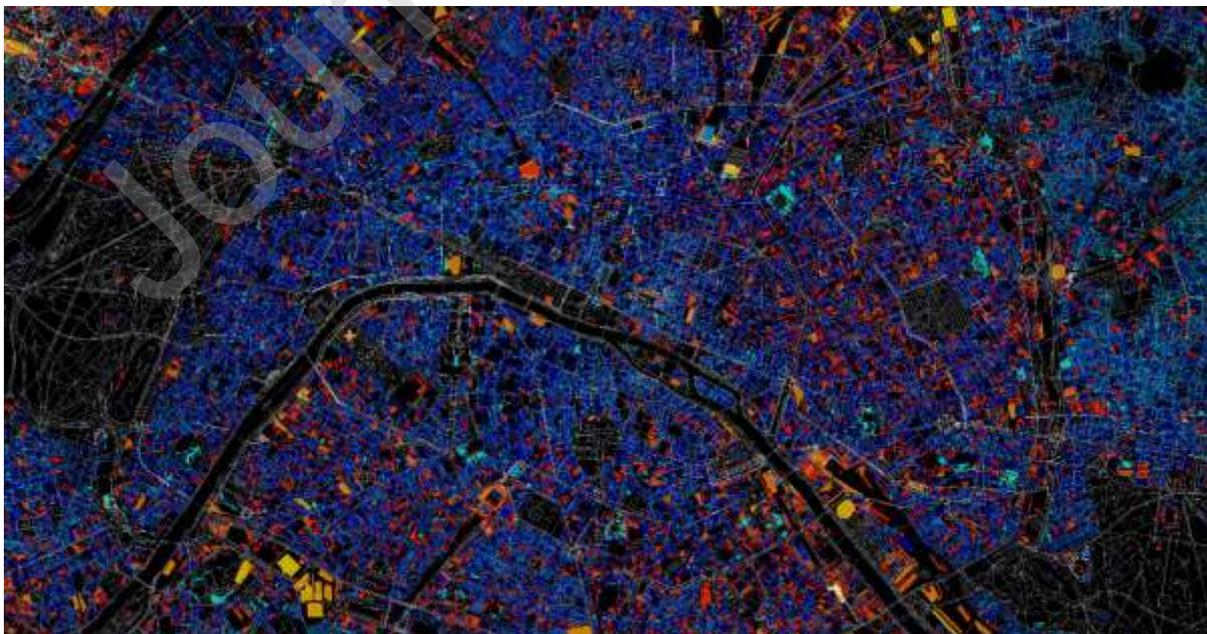
The COVID-19 has led to opposite emotional reactions towards models: some distrust prospective (or retrospective) scenarios, whereas others believe models to hold the answers to any problem. Both are equally wrong. Models only help us grasp our range of options by informing us of potential consequences associated with some choices. Making the choice remains a political decision. In addition, different questions require different models but even a single question will greatly benefit from being tackled using a variety of models since

hypotheses are infinite, and every different angle is enriching. This can offer widely divergent projections that cast doubt on their validity arising primarily from several fundamental misconceptions about modelling: models are only a workable simplification of a real problem. This simplicity is precisely what makes them useful. They are an abstraction of reality allowing us to make projections informed by evidence and hypotheses rather than actual predictions. In a pandemic context, these challenges and limitations should be conveyed with care by modellers when interacting with the public, the media, healthcare professionals, and decision-makers not only for reasons of scientific accuracy but also to uplift scientific literacy. Indeed, “Educate and actively communicate with the public” is in the top ten non-pharmaceutical interventions ranked by models [20].

**Figure 1:** The figure generated using the EPIDEMAP framework [15] shows the buildings in the Paris area as imported from the [OpenStreetMap.org](https://www.openstreetmap.org) database. The colours reflect the building occupancy (colour warmth correlate with building occupancy). Schools and hospitals are treated differently and shown in cyan. This detailed geographical information can greatly inform infectious disease epidemiological modelling.

#### **References:**

- [1] N. M. Ferguson *et al.*, « Report 9 - Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand ». 2020. [En ligne]. Disponible sur:



- <https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-9-impact-of-npis-on-covid-19/>
- [2] W. O. Kermack & A. G. McKendrick, « A contribution to the mathematical theory of epidemics », *Proc R Soc Lond A*, vol. 115, p. 700-721, 1927.
  - [3] Ross, R., & Hudson, H. P. (1917). An application of the theory of probabilities to the study of a priori pathometry.—Part II. *Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character*, 93(650), 212-225.
  - [4] E. Brooks-Pollock, L. Danon, T. Jombart, et L. Pellis, « Modelling that shaped the early COVID-19 pandemic response in the UK », *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 376, n° 1829, p. 20210001, juill. 2021, doi: 10.1098/rstb.2021.0001.
  - [5] A. Cori, N. M. Ferguson, C. Fraser, et S. Cauchemez, « A New Framework and Software to Estimate Time-Varying Reproduction Numbers During Epidemics », *Am. J. Epidemiol.*, vol. 178, n° 9, p. 1505-1512, nov. 2013, doi: 10.1093/aje/kwt133.
  - [6] M. Castro, S. Ares, J. A. Cuesta, et S. Manrubia, « The turning point and end of an expanding epidemic cannot be precisely forecast », *Proc. Natl. Acad. Sci.*, vol. 117, n° 42, p. 26190-26196, oct. 2020, doi: 10.1073/pnas.2007868117.
  - [7] N. Oliver *et al.*, « Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle », *Sci. Adv.*, vol. 6, n° 23, p. eabc0764, juin 2020, doi: 10.1126/sciadv.abc0764.
  - [8] D. A. Rasmussen, E. M. Volz, et K. Koelle, « Phylodynamic Inference for Structured Epidemiological Models », *PLoS Comput. Biol.*, vol. 10, n° 4, p. e1003570, avr. 2014, doi: 10.1371/journal.pcbi.1003570.
  - [9] Bedson, J., Skrip, L. A., Pedi, D., Abramowitz, S., Carter, S., Jalloh, M. F., ... & Althouse, B. M. (2021). A review and agenda for integrated disease models including social and behavioural factors. *Nature human behaviour*, 5(7), 834-846.
  - [10] T. Jombart, A. Cori, X. Didelot, S. Cauchemez, C. Fraser, et N. Ferguson, « Bayesian Reconstruction of Disease Outbreaks by Combining Epidemiologic and Genomic Data », *PLoS Comput. Biol.*, vol. 10, n° 1, p. e1003457, janv. 2014, doi: 10.1371/journal.pcbi.1003457.
  - [11] N. Mideo, S. Alizon, et T. Day, « Linking within- and between-host dynamics in the evolutionary epidemiology of infectious diseases », *Trends Ecol. Evol.*, vol. 23, n° 9, p. 511-517, sept. 2008, doi: 10.1016/j.tree.2008.05.009.
  - [12] Prague M, Alexandre M, Thiébaud R, and Guedj J. Within-host models of SARS-CoV-2: what can it teach us on the biological factors driving virus pathogenesis and transmission ?, *ACCPM*, 2022 (in press).
  - [13] T. Day, S. Alizon, et N. Mideo, « Bridging scales in the evolution of infectious disease life histories: Theory », *Evolution*, vol. 65, n° 12, p. 3448-3461, déc. 2011, doi: 10.1111/j.1558-5646.2011.01394.x.
  - [14] T. Britton et G. Scalia Tomba, « Estimation in emerging epidemics: biases and remedies », *J. R. Soc. Interface*, vol. 16, n° 150, p. 20180670, janv. 2019, doi: 10.1098/rsif.2018.0670.
  - [15] O. Thomine, S. Alizon, C. Boennec, M. Barthelemy, et M. Sofonea, « Emerging dynamics from high-resolution spatial numerical epidemics », *eLife*, vol. 10, p. e71417, oct. 2021, doi: 10.7554/eLife.71417.
  - [16] V. A. Karatayev, M. Anand, et C. T. Bauch, « Local lockdowns outperform global lockdown on the far side of the COVID-19 epidemic curve », *Proc. Natl. Acad. Sci.*, vol. 117, n° 39, p. 24575-24580, sept. 2020, doi: 10.1073/pnas.2014385117.

- [17] E. E. Glennon *et al.*, « Challenges in modeling the emergence of novel pathogens », *Epidemics*, vol. 37, p. 100516, déc. 2021, doi: 10.1016/j.epidem.2021.100516.
- [18] L. M. Espira, A. F. Brouwer, B. A. Han, J. Foufopoulos, et J. N. S. Eisenberg, « Dilution of Epidemic Potential of Environmentally Transmitted Infectious Diseases for Species with Partially Overlapping Habitats », *Am. Nat.*, vol. 199, n° 2, p. E43-E56, févr. 2022, doi: 10.1086/717413.
- [19] T. Day, T. Parsons, A. Lambert, et S. Gandon, « The Price equation and evolutionary epidemiology », *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 375, n° 1797, p. 20190357, avr. 2020, doi: 10.1098/rstb.2019.0357.
- [20] N. Haug *et al.*, « Ranking the effectiveness of worldwide COVID-19 government interventions », *Nat. Hum. Behav.*, vol. 4, n° 12, p. 1303-1312, déc. 2020, doi: 10.1038/s41562-020-01009-0.