Tooling-up for infectious disease transmission modelling

Marc Baguelin[1,2], Graham F. Medley[2], Emily S. Nightingale[2], Kathleen M. O'Reilly[2], Eleanor M. Rees[2], Naomi R. Waterlow[2], Moritz Wagner[2]

(Authors in alphabetical order)
[1] School of Public Health, Infectious Disease Epidemiology, Imperial College London, United Kingdom
[2] Centre for Mathematical Modelling of Infectious Disease, London School of Hygiene and Tropical Medicine, London, UK

Corresponding author: Kathleen M O'Reilly (kathleen.oreilly@lshtm.ac.uk)

"We become what we behold. We shape our tools, and thereafter our tools shape us."
McLuhan, M 1994, 'Understanding media: The extensions of man', MIT Press

Introduction and motivation

The motivations for transmission dynamic modelling are the motivations for this special issue: harnessing available data to inform policy on the control of infectious diseases. The focus is how to get the most valuable and actionable information out of data - models that translate data into evidence for policy. Transmission dynamic modelling is now central to designing and evaluating public health interventions against infectious diseases, especially for intervention programmes. Who should we vaccinate? Should we close schools to limit spread? What vector control option will save the most lives? These questions, if they are to be answered rationally, need a way of predicting the impact in terms of health outcome, e.g. numbers of clinical cases or deaths averted. Policy questions determine the necessary components and the complexity of the model; models need to be simple enough to analyse, but sufficiently detailed to address the questions.

Dynamic models are needed because the essence of infectious diseases is that transmission dynamics are non-linear: incidence is a function of current and past

prevalence. Additionally, the dynamics are typically unobserved, for example asymptomatic infection frequently drives transmission and we rarely directly observe immunity of individuals. Consequently, it is impossible to make a rational decision on, say, vaccination policy simply from recent case numbers. To understand and predict what is happening, and to understand and predict the impact of interventions, the transmission processes must be properly described using a model (May, 1986).

Data are necessary if the models are to be at all applicable: “it is all about the data” (B. T. Grenfell, pers. comm.). Early papers were content to try to understand the observed patterns: Kermack and McKendrick’s finding that “an epidemic, in general, comes to an end, before the susceptible population has been exhausted” was derived from first principles in order to explain the observations from data on a cholera epidemic in Bombay, India (Kermack et al., 1927). As the data become more complete and more detailed, the methods of analysis and fitting the model to data become more complex; and it is this that has substantially changed in the last 20 years of infectious disease modelling. Kermack and McKendrick’s SIR model from 1927 is still the archetypal virus model, but there have been many additions to this simple model, to test the utility of control strategies (Keeling and Gilligan, 2000) and to account for known biases in the data (Finger et al., 2018, as one of many examples).

This field has been rapidly developing, but is maturing as some of the core ideas become fixed. The purpose of this Special Issue on methods for infectious disease analysis is to act as both a marker in time of where the subject is, and as a way into the literature and techniques for those new to the area. Each paper takes one approach or technique (a ‘tool’), typically associated with a particular software platform, and leads the reader from the basic idea into the darker reaches. We have asked the authors to develop specific, relevant examples and we recommend that the reader work through the code and run the examples: it is impossible to learn to swim without getting wet. The first paper describes and provides an overview of the available infectious disease models and a suggested criteria for selecting a specific model (see Box 1).

Choices and trade-offs in inference with infectious disease models (Funk & King)

With the explosion, in the last four decades, of model types, simulation and inference algorithms, and softwares available it can be confusing to choose the appropriate pathway when confronted to modelling a disease related question. In this paper, the authors review the choices available and trade-offs to be made when designing a model and inference method. Such a roadmap should help the infectious disease modeller to tailor their model to optimally exploit the data available in order to answer efficiently the policy question. Examples from influenza transmission in R are provided.

One of the notable features of this field is the exponential growth in confusing acronyms. That SIS and SIR are used both to define a infectious disease model structure (susceptible-infectious-susceptible and susceptible-infectious-removed/recovered) and the methods of inference (sequential importance sampling/resampling) in the pMCMC paper may confuse the uninitiated. Although the field might be maturing on ideas, there is not an agreed terminology or accepted best practice. There are some notable methodological developments that have been omitted from the special issue (for example, history matching (McKinley et al., 2018)), but if the reader is able to grasp the contents of this special issue, they will be in an ideal position to tackle these, and future, developments.

In this introduction, we put the methodological tools presented in this Special Issue in a historical context and briefly describe the contents of each paper. For example, at first it may not be obvious, and perhaps even confusing, why several methods are available using different software to estimate parameters of a transmission model. The intended readership of the Special Issue are early career researchers that may be less familiar with how the field has evolved to where it currently stands. Later career researchers may also find the papers interesting in order to keep up to date with recent methodological developments. Whilst it is far from compulsory, the order in which the papers are presented in this Introduction could be used to determine the order of reading.

The origin and emergence of disease dynamics (1766-1980s)

Bernoulli is usually credited to have developed the first mathematical model for gaining insight into an infectious disease in 1766 (Dietz and Heesterbeek, 2002). Disease dynamics could be seen at this time as a sub-discipline of population dynamics. With this respect three early 20th century studies can be considered as the founding of disease dynamics: Ross's work on malaria transmission (Ross, 1911), McKendrick and Kermack's paper (Kermack et al., 1927) and Reed-Frost's chain binomial model (presented in 1928 but not published at the time) (Lessler and Cummings, 2016).

A major step-change in the use of models to understand infectious diseases was apparent in the late 70's and early 80's where investigation began in full-swing in both the UK and the USA. These developments were motivated by understanding both the chronic effects of persistent parasite infection (for example hookworm), the regularity of viral epidemics such as 'flu and measles, the emergence of epidemics such as HIV/AIDS, and the initiation of large vaccination programmes (Agur et al., 1993; Anderson and May, 1985; Elveback et al., 1976; Grenfell and Anderson, 1989; Longini et al., 1984; May and Anderson, 1987). Whilst most modelling studies emerged from a small number of research groups, there are some notable exceptions (e.g. (Baroyan, 1963) for influenza in the USSR). Although the development of models in order to support policy making was well underway at the end of the 1980s, it was only later, with the emergence of affordable computational power following the advent of personal computers and the increasing availability of digital data, that effective tools to were developed or borrowed from other fields. This period ends about 1993 marked by the publication of Anderson & May (1993) and the series of meetings at the Newton Institute (Grenfell et al., 1995; Medley and Isham, 1996). Heesterbeek's thesis titled R0 and his notable seminar style in which he filled a blackboard with equations, marks the last point at which infectious disease modelling could be done on paper alone (Heesterbeek, 1992). The publications from those meetings are now notable by their lack of statistical innovation, and the apparent paucity of data.

First revolution: Advanced computing and inference (1990s-early 2000s)

Following the early achievements in epidemic modelling, interest grew in capturing the behaviour of more realistic and more complicated dynamic systems, yet this often

resulted in problems which could no longer be solved analytically. Iterative methods therefore became a vital component in the modeller's toolkit, alongside comparisons of model output to specific datasets by minimising the error between model and data, or by using approximations to simplify the set of estimated parameters. These approaches are illustrated within models applied to the use of the expectation maximisation algorithm to estimate parameters from HIV epidemics (Becker, 2016), the 2001 foot-and-mouth outbreak in UK livestock (Keeling et al., 2001; Tildesley et al., 2006), and estimating parameters from measles epidemics prior to the introduction of vaccination (Finkenstädt et al., 2002).

The theory of sampling from an intractable distribution using Markov chains was originally presented in the early fifties, but, having been born from the field of physics, its potential as a tool in epidemic modelling went largely unnoticed. At the time, it was also not very practical for applied use: even with the most advanced tools available to the 1950s analyst, the original Metropolis paper reports that a run of less than 100 iterations took five hours to complete. These methods were brought more mainstream by Geman and Geman 1984, and Gelfand and Smith 1990. These methods, now referred to as Markov chain Monte Carlo (MCMC), to approximate the posterior distribution has provided an incredibly useful and simple iterative method. It is particularly useful inside the Bayesian framework, but also any situation where a probability distribution can be derived by linking the data and the parameters of the model. It provides a way of exploring the parameter space and obtaining representative samples of the parameters compatible with the target distributions. We assume in this special issue that the essentials of MCMC are known. Readers who might want to look for more information in implementing MCMC can refer to (Gilks et al., 1996; van Ravenzwaaij et al., 2018).

While the general principles behind MCMC are simple, the efficient implementation of even the most simple algorithms can be tricky. The arrival of Bayesian analysis Using Gibbs Sampling (BUGS) in the early 90s as a user-friendly software for implementing MCMC brought Bayesian analysis to the mainstream and spurred a revolution in infectious disease modelling. BUGS remains a widely used tool for analysing data including from the perspective of infectious disease models (see Box 2 and Auzenbergs et. al. (2019) in this issue). The use of the BUGS language for fitting

complex models gained popularity in many areas of infectious disease research, but some challenges to using SIR-like models prevented their widespread use, such as the acyclic nature of non-linear models. Until MCMC samplers became easier to write, and researchers skills in computational methods improved enough to write them, a second revolution was required in methodological research in infectious disease epidemiology.

Desirable BUGS in models of infectious diseases (Auzenbergs et al.)

In the last few decades, tools for implementing Bayesian inference have been developing at quite a pace. There is now a wealth of options available, each with its own advantages and disadvantages, and the choice can be intimidating for anyone taking their first steps into the field. This paper introduces the original BUGS software, which remains a key player for many applications despite its first incarnation dating back to the eighties. The authors set out the basic ideas behind the Bayesian approach and introduce the BUGS language for model specification, which has become the basis for several other software packages since. Three examples are presented and discussed in terms of complexity and run time, providing a practical guide for the first-time user as to which options could prove most successful for their problem.

Second Revolution: adding complexity and heterogeneity (early 2000s - present)

As the field of mathematical modelling for infectious diseases has expanded, so has the complexity of models and their relation to data. Specific policy questions have triggered the need for including new layers of heterogeneity and complexity to existing models. A notable example of this type of change is the inclusion of the age structure within transmission models using contact surveys (Mossong et al., 2008). This increase in model complexity has been driven by an improvement in understanding transmission; in the example by Mossong it has been the importance of age-structure in driving transmission of infectious diseases (in this case ‘flu). Without including this model complexity, predictions in the effectiveness of interventions would be incorrect.

This model diversification has required the application of new inference methods, whether completely novel or borrowed from other fields and adapted to epidemiological studies. Developments in statistical methods applied to infectious diseases in this period include accounting for the incubation period to infer transmission from disease data (Donnelly et al., 2003), using data augmentation to account for asymptomatic/undetected transmission (Cauchemez et al., 2006) and likelihood-free parameter estimation (Toni et al., 2009). Fortunately, advances in

computational capabilities, as well as advances in tools and software packages available have allowed for these developments. Infectious disease modellers now require skills in coding (often in several languages) and statistical inference, as well as an ability to manipulate and solve equations. However, with increasing model complexity, statistical inference becomes challenging and sometimes a distraction from the objectives of the analysis (ie. to capture the dynamics of infection through a population). Fortunately, shortcuts have been made increasingly available, such as packages to borrow inference from elsewhere and approximating the full likelihood, which has been helped by the sharing of code to complement equations within the papers. The increasing use of R and python software by researchers has also been transformational, as the software is relatively accessible and the use of libraries of additional code and wrappers to use other languages enable a rapid progression to carry out infectious disease modelling. To this end, several of the papers within this Special Edition make use of R libraries and various wrappers that port other languages into R (Auzenbergs et al., 2019; Chatzilena et al., 2019; Funk and King, 2020), making modelling more accessible.

A particularly useful development has been Approximate Bayesian Computation (ABC) which can be used where the likelihood of a model is intractable. Whilst ideas related to this method have been circulating since at least the 1980s, only in the late 1990s did ABC in its current form start being applied, primarily in the field of genetics (Tavaré et al., 1997). Central to ABCs usefulness is that instead of evaluating the likelihood, specific characteristics of the data are chosen by the researcher, and model output is compared to these characteristics (See Minter and Retkuteb (2019) in this issue and Box 3). The simplicity of the approach enables parameter estimation which would not otherwise be feasible, thus enabling more complex models to be fitted to data. The challenge then becomes the need to identify appropriate summary statistics.

Approximate Bayesian Computation for infectious disease modelling (Minter & Retkuteb)

Many of the more standard fitting techniques such as Markov Chain Monte Carlo (MCMC) require knowledge of the likelihood in order to fit the model to data and estimate the parameters. Approximate Bayesian Computation (ABC) allows modellers to make these estimations without the likelihood, and with great flexibility. This paper is a great introductory step by step guide to ABC, detailing it's usefulness and providing step by step instructions on how to implement it. Once the theory is covered, three examples are used with ABC, highlighting the complexities and importance of the assumptions made.

An interesting example of cross-discipline methodological developments is provided by particle MCMC. Particle filtering (which is the basis for particle MCMC) stems from fluid mechanics, where it has been used since the 1960's. Particle MCMC (pMCMC) combines MCMC methods with Sequential Monte Carlo (SMC) methods and is particularly useful when dealing with 'hidden' states (e.g. true numbers of infected cases in a community) which have to been inferred and are likely to be highly correlated parameters. Its first use in infectious disease modelling were (Dureau et al., 2013) and is becoming an increasingly popular choice as a fitting method for highly complex models where states are not directly observable. However the development of this method has been hindered by the potentially complex mathematical background behind it, in particular in terms of the multiple indexes necessary to write the algorithms. In this issue, Endo et al. (2019)
(see box 4) have designed a tutorial to guide the epidemiological modeller to be able to use this powerful method of inference for their own project.

Introduction to Particle Markov-chain Monte Carlo for disease dynamics modellers (Endo et al.)

A tutorial to guide the infectious disease modeller to use particle Markov-chain Monte Carlo (PMCMC) methods is described. PMCMC are particularly suited to explore models where the states of the model are only indirectly observed and therefore unknown. These hidden states have to be inferred in order to be able to estimate the parameters of the model. Sequential Monte Carlo methods are used to exploit the time structure of these systems and "integrate out" the hidden states thus enabling the exploration of the parameter space in a similar way than with more classical MCMC methods. In this tutorial, PMCMC is presented with a light touch on the theory and some example of R code is developed for the infectious disease dynamics modeller.

Contributing to the delay in complex inference methods entering mainstream usage is the computational skills of the researchers within the field. Many researchers using mathematical modelling do not have computational backgrounds, and therefore availability of reusable software is essential to methods being widely used. Additionally, providing code in a re-usable format was not initially routine and resulted in inefficiencies in carrying out research (in contrast to the practices of sharing code that perhaps stemmed from C++ users in physics). The practice of "reproducible research", including the sharing of code and associated data has become increasingly common and is now frequently requested by research funders (AMS, 2015). The STAN software platform is a great example of reusable software, as it allows

researchers to efficiently implement complex methods such as Hamiltonian Monte Carlo (HMC) and Variational Inference (VI), both of which allow the fitting of more complex models than simpler MCMC algorithms. The paper by Chatzilena et al in this issue ((Chatzilena et al., 2019) and Box 5) gives more details on these methods.

Contemporary statistical inference for infectious disease models using Stan (Chatzilena et al.)

Stan is a platform for statistical modelling, with implementations of various techniques that can be used for fitting statistical models. In particular, Stan is currently the only platform within which the Hamiltonian Monte Carlo (HMC) and variational inference (VI) are available. The authors here focus on the No-U-Turn-Sampler (NUTS) and Automatic Differentiation Variational Inference (ADVI) implementations of HMC and VI provided within Stan. When coupled with the ordinary differential equation solvers also available, Stan is able to fit infectious disease models of increasing complexity. The paper shows how to implement fitting of such models of disease transmission of increasing complexity. The authors then compare NUTS and ADVI and demonstrate trade offs between statistical accuracy and speed of the algorithms.

Discussion and speculation about the future

The tools and approaches presented in this special issue are a mix of the tried-and-tested and those under active development and exploration. The age of an approach should not impact its use, provided the approach remains useful. ‘Fashions’ in infectious disease modelling will rise, only to be replaced by the latest trends, but with time it is possible to see approaches that continue to provide insight, even if the specifics have been updated using modern tools. In this section we offer some thoughts about how methods might develop.

Since the beginning of infectious disease modelling, sweeping, gross assumptions have had to be made in order to make progress. One of the most fundamental is the “closed population”, which only exist in very special circumstances such as boarding schools and isolated populations (Keeling and Grenfell, 1997). But people (and non-human hosts) move. An important problem at the time of writing is the emergence and spread of the novel coronavirus (Covid-19), initially in Wuhan, China (Riou and Althaus, 2020). As well as the need to rapidly understand the natural history of infection of this new virus, it is essential to understand how population movement will impact spread. An important part of the data revolution is the increasing ease with which we can collect geostatistical information; it is increasingly common to use flight data to estimate patterns of emergence and associate an individual with a precise

and temporally changing spatial location, utilising GPS devices and smartphones (Wesolowski et al., 2015). The analysis of infectious disease spread in geographical space and time is an area requiring more attention. The paper by Milton et al. (see Box 6, (Milton et al., 2019)) is different from the others in the issue in that the dynamics are not at the centre of attention, but this approach represents one of our bets for the future. Infectious disease dynamics drive spatio-temporal statistical patterns on an aggregate level, providing an alternative angle for inference when strong assumptions on the transmission process cannot be justified. First considered for measles by Grenfell et al. (2001), methodology for analysing these patterns has been a major growth area and is likely to continue.

Spatial Analysis Made Easy with Linear Regression and Kernels (Milton et al.)

Linear models are commonly the first methodology to be introduced when learning about statistics. They are widely used even though the real world is often far more complex and non-linear. This paper provides an introduction into how such linear models can be extended to non-linear problems using kernel methods, whilst focussing on spatial examples. Furthermore, random Fourier features are introduced which offer a computationally efficient way of implementing kernel methods, so that they can be used for large datasets as well. The paper provides a tutorial-style overview of how these methods can be derived from basic principles including examples and relevant R code. This can be seen as an excellent starting point that encourages the reader to further explore and apply these topics within their own research.

Another safe prediction is that data are only going to get “bigger”. The standardisation and integration of genomic sequencing, georeferencing, and the use of social media will only develop richer and more detailed data on which to build models. All these expansions contribute towards capturing heterogeneity in transmission, which is why their inclusion in many settings will be important. In particular, the inclusion of data-driven dynamic contact networks within infection models is starting to appear, particularly in veterinary disease contexts where the data exist (Orton et al., 2018). The embarrassment for the modeller will be not so much that there is no data to inform parameters (which was the norm 30 years ago), but that there is so much data that its not clear what the model structure should be to accommodate the data. In this context we are likely to see more and more machine learning approaches used to extract and simplify the relationships within the data. Big data and the associated tools to deal with big data might well be the third revolution for infectious disease modelling. With the

increasing complexity of model and data streams it is likely that the field of infectious disease modelling will branch itself into many forms, if it hasn't already.

The interaction between models and policy decisions are also likely to change dramatically in the foreseeable future. The quality of the evidence provided by different models fitted to different datasets is not well defined, and there has been recent interest in developing model ensembles and modelling consortia (Flasche et al., 2016; Hollingsworth and Medley, 2017). Although this solves the problem of basing policy on a single model, it raises the problem of how to combine different model outcomes. Weighting the strength of evidence from each model seems a sensible approach, but choosing the weights is problematic. The majority of the approaches in this issue work in a Bayesian context, where the data and model are combined with prior information to inform parameter estimates and therefore model behaviour. If the public health decision is also put into a Bayesian context, i.e. the model is part of the evidence, then it seems logically natural to treat models as "data".

When the 2001 UK foot-and-mouth outbreak emerged, outputs from infectious diseases models were directly used to inform policy and the decisions made. This was perhaps the start of using real-time epidemic modelling, and with the emergence of SARS, pandemic 'flu (2009), Ebola, Zika virus and now Covid-19 (this is far from an exhaustive list), the appetite of policy makers to use models to aid decision making has not abated. The field needs to keep ahead of the demand, and to this point rather sophisticated methods to infer R0 and other parameters have been developed so that they can be used rapidly and often by researchers with limited expertise (Polonsky et al., 2019). Another aspect of this expanding field is improved assessment of model predictions (Johansson et al., 2019).

In conclusion, since its inception infectious disease modelling has already gone through two revolutions, and it appears that the field is rapidly transitioning into a new one. Can the field remain whole in this new era of Big Data and advanced computing? And if not, where will the splits be? Those who develop inference methods and those who have the infectious disease knowledge to implement them? Or rather into subject-specific domains: those who are able to work with spatial data, those with social data, those with missing data… What does remain clear is that despite extensive

interventions and effort, infectious agents continue to cause widespread morbidity and mortality. Improving on our current systems and fully utilising new data streams remains a daunting task, and it is through collaboration within and outside our field that we will grow into this new revolution, make advancements in global health, and vitally, not misuse data that could cost lives.

Acknowledgements

The authors thank the UK National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Modelling Methodology at Imperial College London in partnership with Public Health England (PHE) for funding (grant HPRU-2012–10080).

References

Agur, Z., Cojocaru, L., Mazor, G., Anderson, R.M., Danon, Y.L., 1993. Pulse mass measles vaccination across age cohorts. Proc. Natl. Acad. Sci. U. S. A. 90, 11698–702. https://doi.org/10.1073/pnas.90.24.11698

AMS, 2015. Reproducibility and reliability of biomedical research: improving research practice. Symposium Report.

Anderson, R., May, R., 1985. Herd-Immunity to Helminth Infection and Implications for Parasite Control. Nature 315, 493–496. https://doi.org/10.1038/315493a0

Auzenbergs, M., Correia-Gomes, C., Economou, T., Lowe, R., O'Reilly, K.M., 2019. Desirable BUGS in models of infectious diseases. Epidemics 29, 100361. https://doi.org/10.1016/j.epidem.2019.100361

Becker, N.G., 2016. Uses of the EM algorithm in the analysis of data on HIV/AIDS and other infectious diseases: Stat. Methods Med. Res. https://doi.org/10.1177/096228029700600103

Cauchemez, S., Boelle, P.Y., Donnelly, C.A., Ferguson, N.M., Thomas, G., Leung, G.M., Hedley, A.J., Anderson, R.M., Valleron, A.J., 2006. Real-time estimates in early detection of SARS. Emerg. Infect. Dis. 12, 110–113.

Chatzilena, A., van Leeuwen, E., Ratmann, O., Baguelin, M., Demiris, N., 2019. Contemporary statistical inference for infectious disease models using Stan. Epidemics 29, 100367. https://doi.org/10.1016/j.epidem.2019.100367

Dietz, K., Heesterbeek, J. a. P., 2002. Daniel Bernoulli's epidemiological model revisited. Math. Biosci. 180, 1–21. https://doi.org/10.1016/s0025-5564(02)00122-0

Donnelly, C.A., Ferguson, N.M., Ghani, A.C., Anderson, R.M., 2003. Extending backcalculation to analyse BSE data. Stat. Methods Med. Res. 12, 177–190. https://doi.org/10.1191/0962280203sm337ra

Dureau, J., Kalogeropoulos, K., Baguelin, M., 2013. Capturing the time-varying drivers of an epidemic using stochastic dynamical systems. Biostatistics 14, 541–555. https://doi.org/10.1093/biostatistics/kxs052

Elveback, L.R., Fox, J.P., Ackerman, E., Langworthy, A., Boyd, M., Gatewood, L., 1976. An influenza simulation model for immunization studies. Am. J. Epidemiol. 103, 152–165. https://doi.org/10.1093/oxfordjournals.aje.a112213

Endo, A., van Leeuwen, E., Baguelin, M., 2019. Introduction to particle Markov-chain Monte Carlo for disease dynamics modellers. Epidemics 29, 100363. https://doi.org/10.1016/j.epidem.2019.100363

Finger, F., Bertuzzo, E., Luquero, F.J., Naibei, N., Touré, B., Allan, M., Porten, K., Lessler, J., Rinaldo, A., Azman, A.S., 2018. The potential impact of case-area targeted interventions in response to cholera outbreaks: A modeling study. PLoS Med. 15, e1002509. https://doi.org/10.1371/journal.pmed.1002509

Finkenstädt, B.F., Bjørnstad, O.N., Grenfell, B.T., 2002. A stochastic model for extinction and recurrence of epidemics: estimation and inference for measles outbreaks. Biostatistics 3, 493–510. https://doi.org/10.1093/biostatistics/3.4.493

Flasche, S., Jit, M., Rodríguez-Barraquer, I., Coudeville, L., Recker, M., Koelle, K., Milne, G., Hladish, T.J., Perkins, T.A., Cummings, D.A.T., Dorigatti, I., Laydon, D.J., España, G., Kelso, J., Longini, I., Lourenco, J., Pearson, C.A.B., Reiner, R.C., Mier-y-Terán-Romero, L., Vannice, K., Ferguson, N., 2016. The Long-Term Safety, Public Health Impact, and Cost-Effectiveness of Routine Vaccination with a Recombinant, Live-Attenuated Dengue Vaccine (Dengvaxia): A Model Comparison Study. PLOS Med. 13, e1002181. https://doi.org/10.1371/journal.pmed.1002181

Funk, S., King, A.A., 2020. Choices and trade-offs in inference with infectious disease models. Epidemics 30, 100383. https://doi.org/10.1016/j.epidem.2019.100383

Gilks, W. R, Richardson, S, Spiegelhalter, D. J., 1996. Markov chain Monte Carlo in practice. Chapman and Hall.

Grenfell, B.T., Anderson, R.M., 1989. Pertussis in England and Wales: an investigation of transmission dynamics and control by mass vaccination 236, 213–252.

Grenfell, B.T., Bjørnstad, O.N., Kappey, J., 2001. Travelling waves and spatial hierarchies in measles epidemics. Nature 414, 716–723. https://doi.org/10.1038/414716a

Grenfell, B.T., Dobson, A.P., Wright, J., 1995. Ecology of infectious diseases in natural populations. Cambridge University Press.

Heesterbeek, J.A.P., 1992. R_0 - PhD thesis. University of Leiden.

Hollingsworth, T.D., Medley, G.F., 2017. Learning from multi-model comparisons: Collaboration leads to insights, but limitations remain. Epidemics 18, 1–3. https://doi.org/10.1016/j.epidem.2017.02.014

Johansson, M.A., Apfeldorf, K.M., Dobson, S., Devita, J., Buczak, A.L., Baugher, B., Moniz, L.J., Bagley, T., Babin, S.M., Guven, E., Yamana, T.K., Shaman, J., Moschou, T., Lothian, N., Lane, A., Osborne, G., Jiang, G., Brooks, L.C., Farrow, D.C., Hyun, S., Tibshirani, R.J., Rosenfeld, R., Lessler, J., Reich, N.G., Cummings, D.A.T., Lauer, S.A., Moore, S.M., Clapham, H.E., Lowe, R., Bailey, T.C., García-Díez, M., Carvalho, M.S., Rodó, X., Sardar, T., Paul, R., Ray, E.L., Sakrejda, K., Brown, A.C., Meng, X., Osoba, O., Vardavas, R., Manheim, D., Moore, M., Rao, D.M., Porco, T.C., Ackley, S., Liu, F., Worden, L., Convertino, M., Liu, Y., Reddy, A., Ortiz, E., Rivero, J., Brito, H., Juarrero, A., Johnson, L.R., Gramacy, R.B., Cohen, J.M., Mordecai, E.A., Murdock, C.C., Rohr, J.R., Ryan, S.J., Stewart-Ibarra, A.M., Weikel, D.P., Jutla, A., Khan, R., Poultney, M., Colwell, R.R., Rivera-García, B., Barker, C.M., Bell, J.E., Biggerstaff, M., Swerdlow, D., Mier-y-Teran-Romero, L., Forshey, B.M., Trtanj, J., Asher, J., Clay, M., Margolis, H.S., Hebbeler, A.M., George, D., Chretien, J.-P., 2019. An open challenge to advance probabilistic forecasting for dengue epidemics. Proc. Natl. Acad. Sci. 116, 24268–24274. https://doi.org/10.1073/pnas.1909865116

Keeling, M. J., Grenfell, B. T., 1997. Disease Extinction and Community Size: Modeling the Persistence of Measles | Science 275, 65–67.

Keeling, M.J., Gilligan, C.A., 2000. Bubonic plague: a metapopulation model of a zoonosis. Proc. Biol. Sci. 267, 2219–2230. https://doi.org/10.1098/rspb.2000.1272

Keeling, M.J., Woolhouse, M.E.J., Shaw, D.J., Matthews, L., Chase-Topping, M., Haydon, D.T., Cornell, S.J., Kappey, J., Wilesmith, J., Grenfell, B.T., 2001. Dynamics of the 2001 UK Foot and Mouth Epidemic: Stochastic Dispersal in a Heterogeneous Landscape. Science 294, 813–817. https://doi.org/10.1126/science.1065973

Kermack, W.O., McKendrick, A.G., Walker, G.T., 1927. A contribution to the mathematical theory of epidemics. Proc. R. Soc. Lond. Ser. Contain. Pap. Math. Phys. Character 115, 700–721. https://doi.org/10.1098/rspa.1927.0118

Lessler, J., Cummings, D.A.T., 2016. Mechanistic Models of Infectious Disease and Their Impact on Public Health. Am. J. Epidemiol. 183, 415–422. https://doi.org/10.1093/aje/kww021

Longini, I.M., Seaholm, S.K., Ackerman, E., Koopman, J.S., Monto, A.S., 1984. Simulation studies of influenza epidemics: assessment of parameter estimation and sensitivity. Int. J. Epidemiol. 13, 496–501. https://doi.org/10.1093/ije/13.4.496

May, R.M., Anderson, R.M., 1987. Transmission dyanmics of HIV infection. Nature 326, 137–142.

McKinley, T.J., Vernon, I., Andrianakis, I., McCreesh, N., Oakley, J.E., Nsubuga, R.N., Goldstein, M., White, R.G., 2018. Approximate Bayesian Computation and Simulation-Based Inference for Complex Stochastic Epidemic Models. Stat. Sci. 33, 4–18. https://doi.org/10.1214/17-STS618

Medley, G.F., V. Isham, 1996. Models for Infectious Human Diseases: Their Structure and Relation to Data, Publications of the Newton Institute.

Milton, P., Coupland, H., Giorgi, E., Bhatt, S., 2019. Spatial analysis made easy with linear regression and kernels. Epidemics 29, 100362. https://doi.org/10.1016/j.epidem.2019.100362

Minter, A., Retkute, R., 2019. Approximate Bayesian Computation for infectious disease modelling. Epidemics 29, 100368. https://doi.org/10.1016/j.epidem.2019.100368

Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Tomba, G.S., Wallinga, J., Heijne, J., Sadkowska-Todys, M., Rosinska, M., Edmunds, W.J., 2008. Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. PLoS Med. 5. https://doi.org/10.1371/journal.pmed.0050074

Orton, R.J., Deason, M., Bessell, P.R., Green, D.M., Kao, R.R., Salvador, L.C.M., 2018. Identifying genotype specific elevated-risk areas and associated herd risk

factors for bovine tuberculosis spread in British cattle. Epidemics 24, 34–42. https://doi.org/10.1016/j.epidem.2018.02.004

Polonsky, J.A., Baidjoe, A., Kamvar, Z.N., Cori, A., Durski, K., Edmunds, W.J., Eggo, R.M., Funk, S., Kaiser, L., Keating, P., de Waroux, O. le P., Marks, M., Moraga, P., Morgan, O., Nouvellet, P., Ratnayake, R., Roberts, C.H., Whitworth, J., Jombart, T., 2019. Outbreak analytics: a developing data science for informing the response to emerging pathogens. Philos. Trans. R. Soc. B-Biol. Sci. 374, 20180276. https://doi.org/10.1098/rstb.2018.0276

R M May, 1986. When two and two do not make four: nonlinear phenomena in ecology. Proc. R. Soc. B Biol. Sci. 228, 241–266.

Riou, J., Althaus, C.L., 2020. Pattern of early human-to-human transmission of Wuhan 2019 novel coronavirus (2019-nCoV), December 2019 to January 2020. Eurosurveillance 25, 7–11. https://doi.org/10.2807/1560-7917.ES.2020.25.4.2000058

Ross, R., 1911. The Prevention of Malaria.

Tavaré, S., Balding, D.J., Griffiths, R.C., Donnelly, P., 1997. Inferring coalescence times from DNA sequence data. Genetics 145, 505–518.

Tildesley, M.J., Savill, N.J., Shaw, D.J., Deardon, R., Brooks, S.P., Woolhouse, M.E.J., Grenfell, B.T., Keeling, M.J., 2006. Optimal reactive vaccination strategies for a foot-and-mouth outbreak in the UK. Nature 440, 83–86. https://doi.org/10.1038/nature04324

Toni, T., Welch, D., Strelkowa, N., Ipsen, A., Stumpf, M.P.H., 2009. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. J. R. Soc. Interface 6, 187–202.

van Ravenzwaaij, D., Cassey, P., Brown, S.D., 2018. A simple introduction to Markov Chain Monte–Carlo sampling. Psychon. Bull. Rev. 25, 143–154. https://doi.org/10.3758/s13423-016-1015-8

Wesolowski, A., Metcalf, C.J.E., Eagle, N., Kombich, J., Grenfell, B.T., Bjørnstad, O.N., Lessler, J., Tatem, A.J., Buckee, C.O., 2015. Quantifying seasonal population fluxes driving rubella transmission dynamics using mobile phone data. Proc. Natl. Acad. Sci. 112, 11114–11119. https://doi.org/10.1073/pnas.1423542112