Research

Authors for correspondence:
Edward L. Ionides
e-mail: ionides@umich.edu
Ning Ning
e-mail: patning@tamu.edu

THE ROYAL SOCIETY
PUBLISHING

Inference on spatiotemporal dynamics for coupled biological populations

Jifan Li[1], Edward L. Ionides[2], Aaron A. King[3,4,5], Mercedes Pascual[5,6] and Ning Ning[1]

[1]Department of Statistics, Texas A&M University, College Station, TX 77843, USA
[2]Department of Statistics, [3]Department of Ecology & Evolutionary Biology, and [4]Center for the Study of Complex Systems, University of Michigan, Ann Arbor, MI 48109, USA
[5]Santa Fe Institute, Santa Fe, NM 87501, USA
[6]Departments of Biology and Environmental Studies, New York University, NY 10012, USA

ELI, 0000-0002-4190-0174; AAK, 0000-0001-6159-3207

Mathematical models in ecology and epidemiology must be consistent with observed data in order to generate reliable knowledge and evidence-based policy. Metapopulation systems, which consist of a network of connected sub-populations, pose technical challenges in statistical inference owing to nonlinear, stochastic interactions. Numerical difficulties encountered in conducting inference can obstruct the core scientific questions concerning the link between the mathematical models and the data. Recently, an algorithm has been proposed that enables computationally tractable likelihood-based inference for high-dimensional partially observed stochastic dynamic models of metapopulation systems. We use this algorithm to build a statistically principled data analysis workflow for metapopulation systems. Via a case study of COVID-19, we show how this workflow addresses the limitations of previous approaches. The COVID-19 pandemic provides a situation where mathematical models and their policy implications are widely visible, and we revisit an influential metapopulation model used to inform basic epidemiological understanding early in the pandemic. Our methods support self-critical data analysis, enabling us to identify and address model weaknesses, leading to a new model with substantially improved statistical fit and parameter identifiability. Our results suggest that the lockdown initiated on 23 January 2020 in China was more effective than previously thought.

1. Introduction

Biological populations may be structured into a collection of densely populated communities separated by sparsely populated regions. The network of communities, which may be cities in a human context, comprise a metapopulation. Motivation for metapopulation modelling arises when some essential feature of the population dynamics cannot be understood from looking at a single location. The dynamics of persistence through local extinctions and reintroductions have been extensively studied in ecology [1,2]. In epidemiology, metapopulation dynamics can be a barrier to the regional elimination and eventual eradication of a pathogen and may determine the successful invasion of a new pathogen or a new strain of an existing pathogen [3]. In other situations, spatiotemporal dynamics may be an unavoidable component of the system under study without being the focus of the investigation [4,5].

Mathematical models for biological systems are used to inform public policy, despite delicate issues in the implementation and interpretation of such models [6]. Indeed, it can be practically impossible to make sense of the nonlinear stochastic interactions driving biological dynamics without

representing them in a model [7,8]. Difficulties arising in developing these models can be categorized as operational or interpretive. Operationally, we seek to fit complex models using statistically valid, reproducible and transparent methods. Interpretive difficulties arise when drawing causal conclusions from fitting models to observational data, giving rise to opportunities for incorrect conclusions owing to missing variables or other forms of model misspecification. When addressing metapopulation dynamics, the network structure creates the operational challenge of working with systems that have high dimensions in addition to nonlinearity, stochasticity and incomplete observability. The modelling assumptions required to describe movement between sub-populations present additional possibilities for model misspecification.

In this article, we develop a data analysis workflow which remedies some limitations of existing methods. Previous approaches had difficulties carrying out likelihood-based inference because existing algorithms required problematic approximations in order to be computationally feasible. Our approach builds on a recent algorithmic development for evaluating and maximizing the likelihood function for metapopulation models. This development enables the use of likelihood-based methods for parameter estimation, confidence interval construction, model selection and the diagnosis of model misspecification. These likelihood-based inference tools are fundamental statistical techniques, known to have favourable theoretical and practical properties, and so they provide a principled framework for data analysis. The inferential toolkit we present has previously been practical only for low-dimensional systems. A concrete demonstration of our workflow, and its relationship to previous approaches, are carried out via an extended example. We then generalize the lessons learned from this example to provide advice for a practical and methodologically principled inference for metapopulation models.

The recent growth in the study of metapopulation dynamics has been driven partly by the COVID-19 pandemic [9–17] and in part by methodological advances facilitating the fitting of metapopulation models to spatiotemporal data. Until the start of this millennium, developing dynamic models with both statistical and scientific justification was a long-standing open problem for even a single population [18]. Over the past two decades, new algorithms [19–22] and software [23–25], together with ever-increasing computational resources, have enabled routine inference for low-dimensional nonlinear partially observed stochastic dynamic systems. However, fundamental algorithmic scalability issues known as the 'curse of dimensionality' lead to difficulties with the high-dimensional systems arising in metapopulation inference. These issues are clearest for Monte Carlo techniques based on importance sampling [26] but are also evident in the need for variational approximations for large Markov chain Monte Carlo (MCMC) calculations [27]. Thus, the data analysis for metapopulation models has lagged behind the analysis of low-dimensional time-series data for biological dynamics. Recent developments enable this gap to be closed, as we demonstrate via a reanalysis of COVID-19, viewed from the context of the ability to draw evidence-based scientific conclusions about the dynamics of the emerging pandemic in January and February 2020.

Biological systems are characterized by nonlinear stochastic dynamics together with incomplete and noisy measurements [18]. We, therefore, focus on the class of partially observed Markov process (POMP) models [28], acknowledging that deterministic models can be conceptually useful but are problematic as statistical explanations of noisy systems [5,29]. The Markov property asserts that the dynamic process has no memory conditional on its current state, which is algorithmically convenient while being scientifically non-restrictive since we can choose what to include in the state. Metapopulation models consider a multivariate system state at each location, and so we require methods tailored for high-dimensional POMP models. Simplifications arise if models and data are limited to binary presence or absence, or a small discrete set of values at each location [2], but we are concerned with situations where time series of abundance data are available, such as case reports for infectious diseases. We focus on two inferential approaches for high-dimensional POMP models, the block particle filter (BPF) and the ensemble Kalman filter (EnKF). Other alternatives are reviewed in electronic supplementary material, section S3.

EnKF was developed in the context of massive geophysical models. It combines an ensemble representation of the latent state with a computationally efficient update rule inspired by the scalable linear Kalman filter, providing an approach with excellent scalability [30,31]. For biological systems, the EnKF was first demonstrated as a computationally convenient tool for compartment models at a single location [32,33]. Subsequently, it has been applied for epidemiological metapopulation inference [9,34]. However, the linearization in the EnKF update rule can be problematic for highly nonlinear systems [31,35]. Furthermore, a linear update rule is not appropriate for small, discrete populations unless the EnKF is embedded within an MCMC algorithm [36]. In contrast, particle filter methods [37] avoid linearization and are directly applicable to discrete and continuous latent states.

For low-dimensional systems, particle filter methods are broadly applicable; they permit consideration of arbitrary nonlinear dynamics and require the model to be specified only via a simulator [23,28]. Particle filters enable statistically efficient use of data, since they provide an evaluation of the likelihood function required for Bayesian or likelihood-based inference, with approximation resulting only from the finite Monte Carlo effort. For high-dimensional systems, scalability considerations demand further approximations since particle filters suffer acutely from the 'curse of dimensionality' [26]. The BPF algorithm modifies the particle filter to achieve scalability by carrying out local resampling on spatial neighbourhoods known as blocks. This avoids the linear update rule used by the EnKF [38]. It is an empirical question whether the different approximations inherent in the EnKF and BPF are successful on metapopulation models, with prior evidence favouring the BPF [35]. In the following example, we demonstrate that the BPF can be effective for a practical metapopulation data analysis. We show that the resulting likelihood-based inference framework provides opportunities for model criticism, leading to rigorous assessment of model fit and improved advice on public policy decisions.

2. Metapopulation analysis of COVID-19 spread in China

We reconsider the influential analysis of COVID-19 by Li *et al.* [9], from early in the pandemic. This analysis provided estimates of transmission parameters and the effect of the lockdown in China using the limited data available at the time. Other teams have fitted models to address similar questions [16,39,40] but the study by Li *et al.* [9] is distinctive for fitting a stochastic mechanistic metapopulation model to extensive spatiotemporal data. The results were published in May 2020 [6], based on reported cases from 10 January to 8 February of that year. The state-of-the-art spatiotemporal analysis was possible on an urgent timescale because the team of researchers had developed their methodology in a sequence of previous situations [32,33,41,42]. The paper is written with attention to reproducibility, and the main results are strengthened by various supporting analyses in an extensive supplement. We show that the approach developed in this article, using algorithms that were not available to Li *et al.* [9], can identify and fix various limitations in their model and inferences. Our goal is not to criticize any specific paper, but rather to build on the timely analysis of Li *et al.* [9] to demonstrate how recently developed techniques provide possibilities to carry out improved data analysis in the future.

For our metapopulation system, the sub-populations are 373 provincial cities in China (meaning cities with administrative responsibility for an entire region) and the data are daily reported COVID-19 cases. COVID-19 dynamics are represented by a susceptible–exposed–asymptomatic–infectious–recovered (SEAIR) epidemic model. Questions of urgent interest early in the pandemic included the relative transmissibility of reported to unreported cases, the fraction of unreported cases and the effect on transmission of movement restrictions imposed on and around 23 January [9]. The model structure is illustrated by the flow diagram in figure 1. Section 4 provides an additional description, with equation (4.1) giving the modelled rate at which susceptible individuals become infected. The complete model specification and estimated parameter values are in electronic supplementary material, section S1.

We consider different model implementations within this structure. Our starting point is model M_1 which is based on the model of [9] and is described in electronic supplementary material, section S1.3. We consider the full dataset, from 10 January to 8 February, with transmission parameters re-estimated following the lockdown on 23 January; these correspond to the periods 1 (10 January to 22 January) and 3 (24 January to 8 February) of [9]. Some minor differences between M_1 and [9] were introduced to enable us to place their model within the general framework of spatiotemporal partially observed continuous-time Markov process models described by Ionides *et al.* [35]. Despite these modifications, simulations from M_1, using the parameters of [9], closely match simulations from the code provided by Li *et al.* [9] (electronic supplementary material, section S1.3). In M_1, the modelled movement rate between cities is proportional to mobility data based on cell phone records from Tencent. However, an inspection of the mobility data reveals that some small cities had no recorded incoming travellers, and therefore, no possibility of introducing SARS-CoV2 within M_1 (or the model of [9]) (electronic supplementary material, section S1.2). This minor limitation formally results in a likelihood of 0 for M_1 (i.e. it is impossible for the simulation model to reproduce the observed spatiotemporal dataset), and hence a log-likelihood of $-\infty$.

We addressed the problematic mobility data in M_1 by adding some additional transportation based on a gravity movement model, as described in electronic supplementary material, section S1.2, giving rise to model M_2. We implemented an additional adjustment between models M_1 and M_2 to align the measurement model with the EnKF inference method presented by Li *et al.* [9]. The EnKF implementation involved specifying a quantity called the observation error variance, defined as a function of the observed cases, to quantify the uncertainty in the measurements. Within the POMP specification, the measurement variance can depend on the latent state but not directly on the observed data. The use of a data-dependent observation error variance is, therefore, inconsistent with the theoretical motivation of the EnKF algorithm as a filter for a POMP model. To interpret the EnKF observation variance chosen by Li *et al.* [9] within the POMP framework, we specified the measurement model for M_2 to have equivalent scaling to the choice of Li *et al.* [9], but with the dependence on the reported cases replaced by dependence on the modelled, but unobserved, exact case count.

Based on a comparison of various nonlinear spatiotemporal filters (electronic supplementary material, figure S7) we evaluated the log-likelihood for M_2 using a BPF (table 1). To account for model overfitting, the number of estimated parameters can be subtracted from the log-likelihood to obtain a comparison equivalent to Akaike's information criterion (AIC) [43]. When the difference in log-likelihood is large compared to the difference in degrees of freedom, the ordering of statistical goodness-of-fit is clear without presenting formal statistical hypothesis tests.

To find out whether this log-likelihood value suggests that the model is satisfactory, we compare it with two simple statistical models: M_3 simply models the daily case report for each city as an independent identically distributed (IID) negative binomial random variable; M_4 adds an autoregressive component to M_3 (see electronic supplementary material, section S2). We see from table 1 that both M_3 and M_4 outperform M_2 by many units of log-likelihood. Likelihood can properly be compared between different models for the same data, with statistical uncertainty in log-likelihood differences arising on the unit scale [44]. When the fit of a mechanistic model is inferior to a simple statistical model, we learn that the mechanistic model has room for improvement as a description of the data, but we do not immediately learn what the deficiency is. The development of methods for formal statistical fitting of mechanistic models has led to an increased understanding of the importance of appropriate modelling of over-dispersed variation in stochastic dynamics [45–47]. We, therefore, hypothesized that the fit of M_2 could be improved by permitting additional dynamic noise.

A standard way to convert a deterministic model, constructed as a system of ordinary differential equations, into a stochastic model is to reinterpret the rates of the deterministic system as rates of a Markov chain [48]. This places limits on the mean–variance relationship of the resulting stochastic model [49]. Models allowing greater variability than permitted by this construction are said to be over-dispersed. We added multiplicative white noise to the transmission rate, following the approach of [28,45], giving rise to model M_5. We fitted the model using an iterated BPF to maximize the likelihood [50,51]. The block
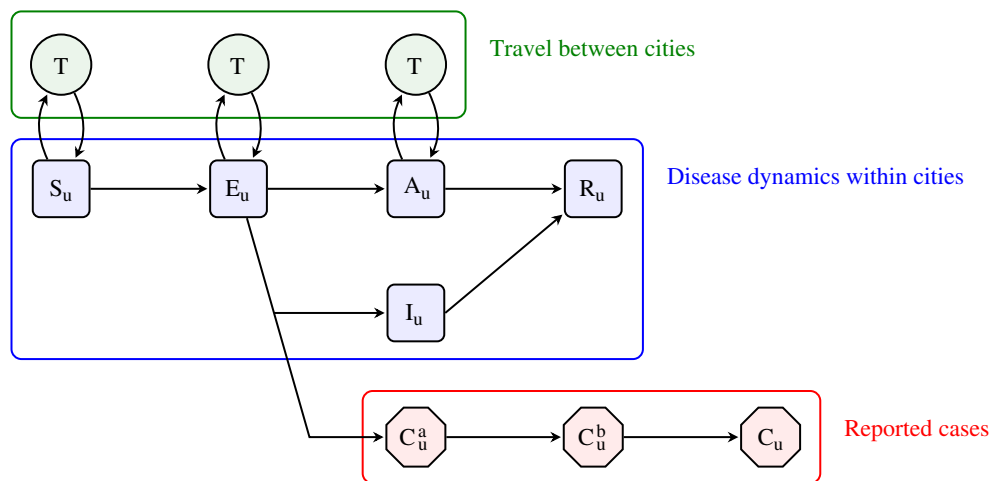
Figure 1. A flow diagram for the SEAIR metapopulation model. Each individual in city u is a member of exactly one of the square blue compartments. Individuals entering the reportable infectious compartment, I_u for city u, are simultaneously included in the delayed reporting process compartment, C_u^a. Upon arrival at the final reporting compartment, C_u, the individual is included in the case report for city u. Individuals in A_u are not reportable and transmit at a reduced rate. The movement of individuals between cities occurs by transport to and from a transport compartment, T. The number of individuals moving between each pair of cities is based on the 2018 data from Tencent. The movement is modelled only for susceptible, exposed and undetected infections.

Table 1. Model comparisons by log-likelihood, evaluated by a block particle filter. The degrees of freedom (d.f.) is the number of estimated parameters.

model	log-likelihood	d.f.	description
M_1	$-\infty$	11	SEAIR model using the parameter values and mobility data of [9]
M_2	$-14\,985.0$	11	adjusted mobility and measurement in M_1
M_3	$-11\,257.9$	374	independent identically distributed negative binomial
M_4	$-10\,825.3$	375	autoregressive negative binomial
M_5	-9088.2	15	adding overdispersed dynamics to M_2 and refitting
M_6	-9116.5	13	latent and infectious durations unchanged by lockdown in M_5.

filter approximation has also proven helpful for spatiotemporal inference when using alternatives to particle filtering and alternatives to maximization by iterated filtering [47]. In the current context, the BPF was found to be more effective for likelihood evaluation than a test suite of alternative filters including the EnKf (electronic supplementary material, figure S7). The iterated BPF (IBPF) maximizes the BPF likelihood using an iterated filtering algorithm [22] adapted to the structure of a BPF.

Table 1 shows that model M_5 outperforms simple statistical benchmarks, obtaining a competitive likelihood with relatively few parameters. From a statistical perspective, M_5 is, therefore, an adequate statistical description of the data. However, some parameters of M_5 were weakly identified by the data, especially in the pre-lockdown time interval within which there were relatively few reported cases (electronic supplementary material, section S6). When the evidence about the model parameters in the data is weak, the ambiguity may be resolved by other, unmodelled and poorly understood, aspects of the data. This risks leading to undesirable situations where important conclusions about questions of interest could be driven by the weaknesses of the model rather than its strengths. In electronic supplementary material, section S6, we show how the flexibility of M_5 can be used to obtain a high likelihood via an implausibly long estimated duration of infection during the pre-lockdown period, with the estimate suddenly reducing after lockdown. We resolved this issue by constraining the latent and infectious periods to be the same before and after lockdown, leading to model M_6. A time-varying effective infectious period could be appropriate owing to changes in quarantine practices, so we let the data assess whether a constant infectious period is plausible in the context of this model. The additional constraints of M_6 lead to a small loss of likelihood compared to M_5, but the fit remains competitive compared to the benchmark models, and the stronger identifiability facilitates the interpretation of estimated parameters. Calculating the maximized log-likelihood for each model in table 1 requires extensive computation to produce a single number which contains essentially all the information about the statistical fit of the model. However, deeper investigation is required to understand what characteristics of the models and data cause the differences in these numbers, and the practical consequences of the numerical results. As a starting point, figure 2 plots the data next to simulations from models M_1 and M_6. Visually, the comparison confirms M_6 as a reasonable representation of the data. Both M_1 and M_6 overestimate cases before day 14, but the context of rapidly increasing awareness and growing diagnostic capabilities is hard to quantify. A visualization of the entire spatiotemporal dataset cannot provide high resolution on specific features of interest, and so this plot is complemented by various other graphical representations in the supplement.

Parameter values for models M_1, M_5 and M_6 are reported in electronic supplementary material, table S1. Here, we discuss the estimated basic reproductive number (i.e. the expected number of secondary infections from one index case in a fully
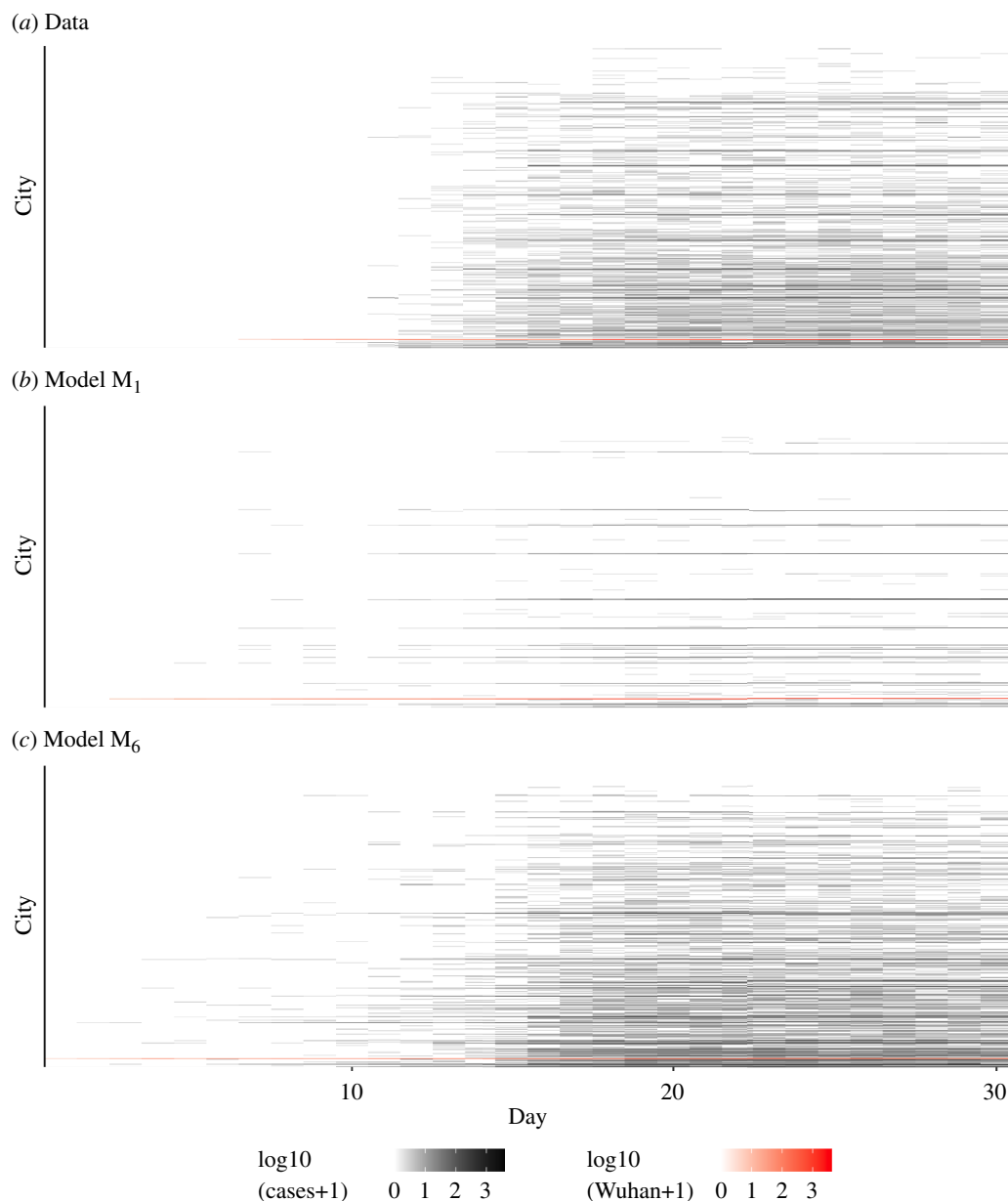
Figure 2. Daily case report time series for 373 cities: (*a*) the real data; (*b*) a simulation from model M_1; (*c*) a simulation from model M_6. Within each panel, cities are ordered by population, the largest being on the bottom row.

susceptible population), denoted by R_0^{be} and R_0^{af} before and after the 23 January lockdown. R_0 is calculated by the formula in equation (4.2). Our estimates for model M_6 are $R_0^{be} = 3.51$ with confidence interval (CI) (3.31, 3.72) and $R_0^{af} = 0.70$ with CI (0.65, 0.77), where the estimates and their associated 95% CIs are obtained by profile likelihood (electronic supplementary material, section S7). This implies that the Chinese government's non-pharmaceutical interventions instituted on and around 23 January reduced R_0 by a factor of 5.0. In contrast, the estimates of [9] are $\tilde{R}_0^{be} = 2.38$ with CI (2.03, 2.77) and $\tilde{R}_0^{af} = 0.98$ with CI (0.83, 1.16), implying reduction by a factor of 2.4. For comparison, interventions implemented across a panel of 41 countries (34 European) were estimated to reduce R_0 by a factor of 4.3 with CI (2.9, 6.7) [52]. Our estimate for R_0 before lockdown is towards the high end of previous estimates based on data up to February 2020, reviewed by Park *et al.* [53]. An alternative metapopulation analysis of the pre-lockdown China data, with a deterministic transmission model, obtained an R_0 estimate of 3.11 with CI (2.39, 4.13) [54]. Our R_0 estimate is consistent with pre-lockdown estimates from other locations, such as New York City, for models that include asymptomatics [55].

The likelihood-based confidence intervals for M_6 are narrower than the intervals from [9]. However, M_6 fits two fewer parameters than M_5, and the latter is more directly comparable to the model of [9]. For M_5, the likelihood-based analysis leads to wide CIs for some parameters, revealing the weakly identified parameters.

Our model inherits the property of [9] that infections arising during the pre-lockdown period will generally be reported during the lockdown, owing to the reporting delay modelled as a distributed delay with a mean of 9 days pre-lockdown and 6 days post-lockdown. Thus, the model is permitted to explain the data by inferring rapid, unreported spread prior to 23 January. Despite this shared constraint on the form of the model, conclusions of our analysis differ from [9]. Beyond the estimates of R_0, a notable difference is that we find the estimated transmissibility of observed cases is close to that of unobserved cases, especially before lockdown (electronic supplementary material, table S1).

If a mechanistic model has a likelihood competitive with statistical benchmarks, it is anticipated to have simulations that are qualitatively comparable to the data. Since the model specification is inevitably imperfect and is accounted for in the model fitting by noise processes, we expect simulations from the fitted model to have somewhat more stochastic variation than the data. In contrast, models that are structurally unable to provide sufficient variability to explain the data must give rise to simulations with too little stochasticity (as shown in figure 2). Models that have simulations with implausibly little variability give rise to claims of excessive confidence about the uncertainty surrounding estimated parameters. This phenomenon may be clearest when CIs are calculated using the parametric bootstrap approach, involving re-estimation of parameters using artificial datasets simulated from a fitted model. However, it also applies to classical CI and Bayesian credible interval constructions. Thus, CIs from mechanistic models that outperform statistical benchmarks are anticipated to be conservative, whereas CIs from models with insufficient variability to explain the data are generally anti-conservative. Requiring model likelihoods to be comparable to statistical benchmarks, therefore, improves the credibility of uncertainty intervals as well as improving the accuracy of point estimates.

3. Discussion

Advances in statistical methodology will drive an increase in the number of spatiotemporal models fitted to epidemiological data. Our research demonstrates that techniques proven effective in low-dimensional systems, such as population dynamics at one or two locations, can be extended to address larger metapopulation systems. This extension allows us to leverage well-established best practices from time-series analysis, leading to a statistically principled approach. This approach enables us to identify and rectify model limitations that might otherwise remain undetected. A failure to address these weaknesses can lead to issues of irreproducibility and the provision of suboptimal policy recommendations when developing models for complex dynamic systems [6,56]. Principles of good data analysis for population dynamics are presumably similar to the general principles of data science [57] but require some adaptation to the specific situation. Here, we build on earlier analyses [6,56,57], by demonstrating the feasibility and desirability of metapopulation analysis meeting the specific set of criteria outlined below. Our implementation of these criteria is summarized in the data analysis workflow shown in figure 3.

— *Likelihood-based statistical inference*: a model, in conjunction with data, defines a likelihood function that quantifies the goodness of fit of the model and the data for each parameter value. For mechanistic models, it is usually impossible to write down the likelihood explicitly, but it still exists implicitly. Likelihood-based methods can extract all available information in the data about model parameters [44]. Log-likelihood is also a proper scoring rule for comparing probabilistic forecasts [58] and, therefore, provides a sensitive objective tool for model selection and identification of model misspecification. Whereas cross-validation and out-of-sample fit are standard benchmarks in machine learning settings [57], likelihood is better suited to situations with relatively small, spatiotemporally structured datasets. Likelihood-based inferences via particle filters have been inaccessible for metapopulation models owing to the 'curse of dimensionality' [26]. However, BPF methods can be effective on metapopulation models, as demonstrated in this paper and previously [5,35,51]. All high-dimensional nonlinear filters entail numerical approximation, and these can be assessed by comparing predictive skill (i.e. the estimated log-likelihood) between different filters. The EnKF provides a suitable point of comparison, since it has excellent scalability properties, modest capability to handle nonlinearities, and has been demonstrated on various epidemiological systems [13,16,32–34,41,42].

— *Statistical benchmarks*: the challenge of fitting intricate nonlinear models to extensive datasets makes it difficult for researchers to evaluate the limitations of their models and methods. Readers also can struggle to determine whether the proposed model has been adequately tested. It is, therefore, advisable to incorporate benchmarks for evaluating model performance in comparison with relatively simple statistical models [45]. This approach helps determine whether complex models provide a satisfactory level of explanatory power. In the first instance, these benchmarks can be applied to the entire dataset; subsequent analysis can focus on dissecting the contributions from various subsets of the data to gain a comprehensive understanding of which parts of the data drive the overall assessment. The likelihood provides a suitable quantity for comparison between different models for the same data [44]. If we find a simple statistical model with a log-likelihood many units higher than a mechanistic model, we have discovered that the mechanistic model is unable to explain some substantial aspects of the data. At the very least, such a discrepancy should be identified and discussed.

— *Residual analysis*: introductory statistics classes, when covering linear regression, emphasize that a careful and complete data analysis involves examining deviations from the fitted model [59]. This is typically achieved by plotting residuals, a suitably rescaled measure of disparities between each observation and its corresponding fitted value. A relevant measure of residual in the current context is the log-likelihood anomaly, defined as the discrepancy between the mechanistic fit and a benchmark for components of the likelihood at each observation. Electronic supplementary material, section S8 describes how these diagnostic tools were used for developing and evaluating model M_6.

— *Uncertainty*: reliable conclusions should be robust to plausible variations in data, models and algorithms [57]. Standard statistical methods provide measures of uncertainty and the validity of these measures depends critically on the statistical validity of the model. Appropriate modelling of overdispersion can be critical for an accurate assessment of uncertainty for dynamic models [28,45–47].

In addition to the methodological points listed above, reproducibility and extendability are important practical considerations. Observational studies are not generally replicable in an experimental sense. However, the numerical conclusions should be readily reproducible from the observations. A substantial part of the value of a computational model is that it permits *in silico* experimentation of the modelled system. The authors should build and share a computational environment that not only
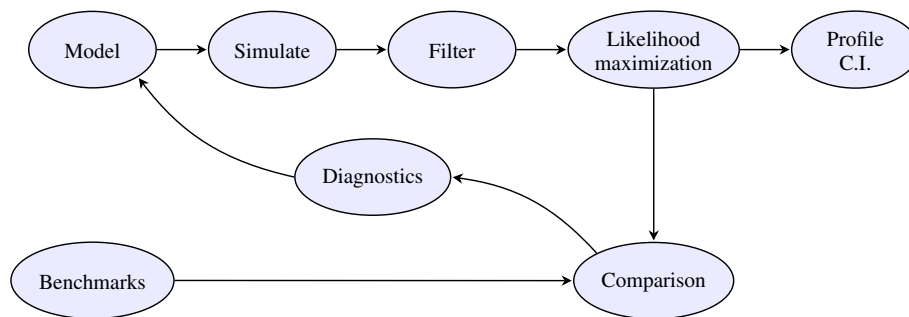
Figure 3. A workflow for likelihood-based data analysis of partially observed spatiotemporal systems. Electronic supplementary material, section S10 provides an additional description of the steps in this workflow.

reproduces published numbers but also facilitates future *in silico* experimentation. Subsequent research should readily be able to challenge the assumptions of the model in light of subsequent data. In practice, this requires that the scientists provide a free, open-source software environment within which the published analysis can readily be reproduced, modified and extended [5,60]. The development of a principled data analysis environment assists the researchers to explore their own models and data, and this environment should be shared as part of the publication process. In practice, this involves encapsulating data analysis within a software package that immerses the user in a documented environment where the models, methods and data used for the article can be readily experimented with. Trustworthy data analysis should be supported by unit testing and documentation, and the quality of this support should be one of the considerations in evaluating the data analysis. In other words, the article presenting the research should be part of a compendium [60]. The compendium for this article comprises the article source code, at https://github.com/jifanli/metapop_article, together with the software environment for the data analysis, provided by the R package at https://github.com/jifanli/metapoppkg.

In conclusion, the study of metapopulation dynamics will continue to benefit from advances in algorithms, software and data analysis methodologies. The models should undergo critical scrutiny to delineate their strengths and weaknesses, following evaluation procedures such as we have described in this paper. With due care, these models can unearth limitations in existing knowledge, investigate hypotheses that may extend our knowledge, and furnish us with valuable predictive tools.

4. Methods

4.1. Data

COVID-19 case reports, city population counts and the time-varying matrix of movement between cities were taken from [9]. Some erroneous numbers, revealed by our log-likelihood anomaly analysis, were subsequently modified as described in electronic supplementary material, section S1.

4.2. Model

All the mechanistic models under consideration have an SEAIR structure, as described in figure 1. Section S1, electronic supplementary material, provides a full mathematical representation of the SEAIR model. Briefly, the force of infection on the susceptible individuals for city u owing to symptomatic and asymptomatic individuals in city u is given by

$$\mu_{S_u E_u} = \beta \left(\frac{I_u(t) + \mu A_u(t)}{P_u(t)} \right) d\Gamma_u/dt, \tag{4.1}$$

where β is the transmission rate, μ is the relative transmissibility of asymptomatic cases and P_u is the city population. The gamma white noise process, $d\Gamma_u/dt$, allows for stochastic variation in the transmission rate [28]. The rate at which individuals move between each pair of cities is defined by a time-varying matrix based on high-resolution Tencent data from 2018, as described in electronic supplementary material, section S1.2. The basic reproductive number is given by

$$R_0 = (\alpha + (1 - \alpha)\mu)D\beta, \tag{4.2}$$

where D is the mean infectious period and α is the fraction of cases severe enough to be reported.

4.3. Likelihood evaluation and maximization

The log-likelihood for the SpatPOMP models was calculated using the BPF. This log-likelihood was then maximized using an IBPF [50,51]. A diagram representing the IBPF algorithm is shown in figure 4. The inner loop, $j = 1, \ldots, J$, corresponds to the BPF applied to an extension of the model where parameters are perturbed by random noise, allowing the resampling step to provide Darwinian natural selection among the population of particles which favours parameter values consistent with the data. The outer loop, $m = 1, \ldots, M$, iterates this BPF procedure while decreasing the magnitude of the perturbations, which
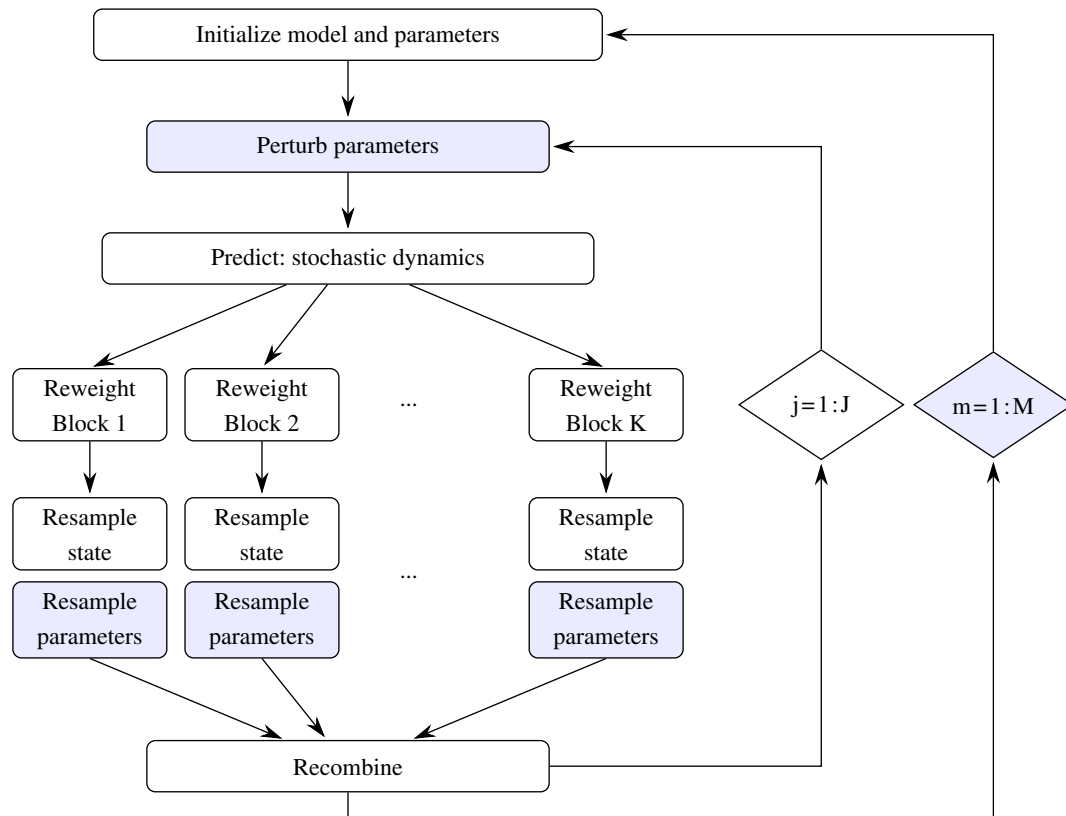
Figure 4. A flow diagram for an iterated block particle filter. The inner loop is a block particle filter, where the high-dimensional latent state is divided into K blocks. The J particles each contain a value for the entire latent state. The outer loop, together with the parameter perturbation and parameter resampling steps, comprises M filtering iterations for maximum likelihood parameter estimation.

is theoretically guaranteed to guide the parameters towards a neighbourhood of the maximum likelihood estimate [22,51]. Further discussion of BPF and IBPF is in electronic supplementary material, section S4. Using this maximization procedure, we constructed confidence intervals by profile likelihood, employing Monte Carlo adjusted profiles [61,62] to correct for Monte Carlo variability. A single likelihood search took 8.5 h on a single computing core, using $J = 1000$ particles and $M = 50$ search iterations. Successful searches were continued for up to eight rounds of refinement until no further gain in likelihood was available. Multiple cores on a Linux cluster were used for Monte Carlo replication and parallel calculation of likelihood profiles.

4.4. Model criticism

A negative binomial autoregressive model was used to provide a non-mechanistic benchmark log-likelihood, as described in electronic supplementary material, section S2. This model was also used to construct benchmark conditional log-likelihoods for each separate observation. These, differenced from the corresponding SEAIR log-likelihoods, were used to define anomalies. The anomalies were explored to identify data points that were poorly explained by the model (electronic supplementary material, section S8). In the preliminary data analysis, these anomalies helped to identify some errors in the data, which were subsequently corrected (electronic supplementary material, section S1.3).

4.5. Software environment

The numerical work was carried out in R [63]. Models and data analysis methodology were developed in an R package, metapoppkg, which is additionally designed to assist the reproducibility and extendability of our results. Models in metapoppkg are implemented using spatPomp [64], which provides a general representation of SpatPOMP models extending the POMP model representation in pomp [23].

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

References

1. Hanski I. 1998 Metapopulation dynamics. *Nat. New Biol.* **396**, 41–49. (doi:10.1038/23876)
2. MacKenzie DI, Nichols JD, Seamans ME, Gutiérrez R. 2009 Modeling species occurrence dynamics with multiple states and imperfect detection. *Ecology* **90**, 823–835. (doi:10.1890/08-0141.1)
3. Metcalf CJE, Andriamandimby SF, Baker RE, Glennon EE, Hampson K, Hollingsworth TD, Klepac P, Wesolowski A. 2021 Challenges in evaluating risks and policy options around endemic establishment or elimination of novel pathogens. *Epidemics* **37**, 100507. (doi:10.1016/j.epidem.2021.100507)
4. Zhang B *et al*. 2022 Mechanisms for the circulation of influenza A(H3N2) in China: a spatiotemporal modelling study. *PLoS Pathog.* **18**, e1011046. (doi:10.1371/journal.ppat.1011046)
5. Wheeler J, Rosengart AL, Jiang Z, Tan K, Treutle N, Ionides EL. 2024 Informing policy via dynamic models: cholera in Haiti. *PLoS Comput. Biol.* **20**, e1012032. (doi:10.1371/journal.pcbi.1012032)
6. Saltelli A *et al*. 2020 Five ways to ensure that models serve society: a manifesto. *Nat. New Biol.* **582**, 482–484. (doi:10.1038/d41586-020-01812-9)
7. McCabe R, Donnelly CA. 2021 Disease transmission and control modelling at the science–policy interface. *Interface Focus* **11**, 20210013. (doi:10.1098/rsfs.2021.0013)
8. Lofgren ET *et al*. 2014 Mathematical models: a key tool for outbreak response. *Proc. Natl Acad. Sci. USA* **111**, 18095–18096. (doi:10.1073/pnas.1421551111)
9. Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, Shaman J. 2020 Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **368**, 489–493. (doi:10.1126/science.abb3221)
10. Wu JT, Leung K, Leung GM. 2020 Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet* **395**, 689–697. (doi:10.1016/S0140-6736(20)30260-9)
11. Wang P, Zheng X, Liu H. 2022 Simulation and forecasting models of COVID-19 taking into account spatio-temporal dynamic characteristics: a review. *Front. Pub Health.* **10**, 1033432. (doi:10.3389/fpubh.2022.1033432)
12. Prieto K, Chávez-Hernández MV, Romero-Leiton JP. 2022 On mobility trends analysis of COVID-19 dissemination in Mexico City. *PLoS One* **17**, e0263367. (doi:10.1371/journal.pone.0263367)
13. Cascante-Vega J, Cordovez JM, Santos-Vega M. 2022 Estimating and forecasting the burden and spread of Colombia's SARS-CoV2 first wave. *Sci. Rep.* **12**, 1–12. (doi:10.1038/s41598-022-15514-x)
14. Pizzuti C, Socievole A, Prasse B, Van Mieghem P. 2020 Network-based prediction of COVID-19 epidemic spreading in Italy. *Appl. Netw. Sci.* **5**, 1–22. (doi:10.1007/s41109-020-00333-8)
15. Alleman TW, Vergeynst J, De Visscher L, Rollier M, Torfs E, Nopens I, Baetens JM, Belgian Collaborative Group on COVID-19 Hospital Surveillance. 2021 Assessing the effects of non-pharmaceutical interventions on SARS-Cov-2 transmission in Belgium by means of an extended SEIQRD model and public mobility data. *Epidemics* **37**, 100505. (doi:10.1016/j.epidem.2021.100505)
16. Yang W *et al*. 2021 Estimating the infection-fatality risk of SARS-CoV-2 in New York City during the spring 2020 pandemic wave: a model-based analysis. *Lancet Infect. Dis.* **21**, 203–212. (doi:10.1016/S1473-3099(20)30769-6)
17. Engebretsen S *et al*. 2023 A real-time regional model for COVID-19: probabilistic situational awareness and forecasting. *PLoS Comput. Biol.* **19**, e1010860. (doi:10.1371/journal.pcbi.1010860)
18. Bjørnstad ON, Grenfell BT. 2001 Noisy clockwork: time series analysis of population fluctuations in animals. *Science* **293**, 638–643. (doi:10.1126/science.1062226)
19. Ionides EL, Bretó C, King AA. 2006 Inference for nonlinear dynamical systems. *Proc. Natl Acad. Sci. USA* **103**, 18438–18443. (doi:10.1073/pnas.0603181103)
20. Toni T, Welch D, Strelkowa N, Ipsen A, Stumpf MP. 2009 Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J. R. Soc. Interface* **6**, 187–202. (doi:10.1098/rsif.2008.0172)
21. Andrieu C, Doucet A, Holenstein R. 2010 Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. B* **72**, 269–342. (doi:10.1111/j.1467-9868.2009.00736.x)
22. Ionides EL, Nguyen D, Atchadé Y, Stoev S, King AA. 2015 Inference for dynamic and latent variable models via iterated, perturbed Bayes maps. *Proc. Natl Acad. Sci. USA* **112**, 719–724. (doi:10.1073/pnas.1410597112)
23. King AA, Nguyen D, Ionides EL. 2016 Statistical inference for partially observed Markov processes via the R package pomp. *J. Stat. Softw.* **69**, 1–43. (doi:/10.48550/arXiv.1509.00503)
24. Kristensen K, Nielsen A, Berg CW, Skaug H, Bell BM. 2016 TMB: automatic differentiation and Laplace approximation. *J. Stat. Softw.* **70**, 5. (doi:10.18637/jss.v070.i05)
25. de Valpine P, Turek D, Paciorek CJ, Anderson-Bergman C, Temple Lang D, Bodik R. 2017 Programming with models: writing statistical algorithms for general model structures with NIMBLE. *J. Comput. Graph. Stat.* **26**, 403–413. (doi:10.1080/10618600.2016.1172487)
26. Bengtsson T, Bickel P, Li B. 2008 Curse-of-dimensionality revisited: collapse of the particle filter in very large scale systems. In *Probability and statistics: essays in honor of David A. Freedman* (eds T Speed, D Nolan), pp. 316–334. Beachwood, OH: Institute of mathematical statistics.
27. Blei DM, Kucukelbir A, McAuliffe JD. 2017 Variational inference: a review for statisticians. *J. Am. Stat. Assoc.* **112**, 859–877. (doi:10.1080/01621459.2017.1285773)
28. Bretó C, He D, Ionides EL, King AA. 2009 Time series analysis via mechanistic models. *Ann. Appl. Stat.* **3**, 319–348. (doi:10.1214/08-AOAS201)
29. King AA, Domenech de Cellès M, Magpantay FMG, Rohani P. 2015 Avoidable errors in the modelling of outbreaks of emerging pathogens, with special reference to Ebola. *Proc. R. Soc. B* **282**, 20150347. (doi:10.1098/rspb.2015.0347)
30. Evensen G. 2009 Data assimilation. In *Data assimilation: the ensemble Kalman filter*. Berlin, Germany: Springer Science & Business Media. (doi:10.1007/978-3-642-03711-5). See https://link.springer.com/10.1007/978-3-642-03711-5.

31. Evensen G, Vossepoel FC, van Leeuwen PJ. 2022 *Data assimilation fundamentals: a unified formulation of the state and parameter estimation problem*. Cham, Switzerland: Springer Nature. (doi:10.1007/978-3-030-96709-3). See https://link.springer.com/10.1007/978-3-030-96709-3.

32. Shaman J, Karspeck A. 2012 Forecasting seasonal outbreaks of influenza. *Proc. Natl Acad. Sci. USA* **109**, 20425–20430. (doi:10.1073/pnas.1208772109)

33. Yang W, Karspeck A, Shaman J. 2014 Comparison of filtering methods for the modeling and retrospective forecasting of influenza epidemics. *PLoS Comput. Biol.* **10**, e1003583. (doi:10.1371/journal.pcbi.1003583)

34. Kramer SC, Pei S, Shaman J. 2020 Forecasting influenza in Europe using a metapopulation model incorporating cross-border commuting and air travel. *PLoS Comput. Biol.* **16**, e1008233. (doi:10.1371/journal.pcbi.1008233)

35. Ionides EL, Asfaw K, Park J, King AA. 2023 Bagged filters for partially observed interacting systems. *J. Am. Stat. Assoc.* **118**, 1078–1089. (doi:10.1080/01621459.2021.1974867)

36. Katzfuss M, Stroud JR, Wikle CK. 2020 Ensemble Kalman methods for high-dimensional hierarchical dynamic space–time models. *J. Am. Stat. Assoc.* **115**, 866–885. (doi:10.1080/01621459.2019.1592753)

37. Doucet A, Johansen A. 2011 A tutorial on particle filtering and smoothing: fifteen years later. In *Oxford handbook of nonlinear filtering* (eds D Crisan, B Rozovsky). Oxford, UK: Oxford University Press.

38. Rebeschini P, van Handel R. 2015 Can local particle filters beat the curse of Dimensionality? *Ann. Appl. Probab.* **25**, 2809–2866. (doi:10.1214/14-AAP1061)

39. Kraemer MU *et al*. 2020 The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* **368**, 493–497. (doi:10.1126/science.abb4218)

40. Brett TS, Bansal S, Rohani P. 2023 Charting the spatial dynamics of early SARS-CoV-2 transmission in Washington state. *PLoS Comput. Biol.* **19**, e1011263. (doi:10.1371/journal.pcbi.1011263)

41. Yang W, Lipsitch M, Shaman J. 2015 Inference of seasonal and pandemic influenza transmission dynamics. *Proc. Natl Acad. Sci. USA* **112**, 2723–2728. (doi:10.1073/pnas.1415012112)

42. Pei S, Kandula S, Yang W, Shaman J. 2018 Forecasting the spatial transmission of influenza in the United States. *Proc. Natl Acad. Sci. USA* **115**, 2752–2757. (doi:10.1073/pnas.1708856115)

43. Burnham KP, Anderson DR. 2002 *Model selection and inference: a practical information-theoretic approach*, 2nd ed. New York, NY: Springer-Verlag.

44. Pawitan Y. 2001 *All likelihood: statistical modelling and inference using likelihood*. Oxford, UK: Oxford University Press. (doi:10.1093/oso/9780198507659.001.0001)

45. He D, Ionides EL, King AA. 2010 Plug-and-play inference for disease dynamics: measles in large and small towns as a case study. *J. R. Soc. Interface* **7**, 271–283. (doi:10.1098/rsif.2009.0151)

46. Stocks T, Britton T, Höhle M. 2020 Model selection and parameter estimation for dynamic epidemic models via iterated filtering: application to rotavirus in Germany. *Biostatistics* **21**, 400–416. (doi:10.1093/biostatistics/kxy057)

47. Whitehouse M, Whiteley N, Rimella L. 2023 Consistent and fast inference in compartmental models of epidemics using Poisson approximate likelihoods. *J. R. Stat. Soc. B* **85**, 1173–1203. (doi:10.1093/jrsssb/qkad065)

48. Keeling M, Rohani P. 2008 *Modeling infectious diseases in humans and animals*. Princeton, NJ: Princeton University Press. (doi:10.1515/9781400841035). See https://www.degruyter.com/document/doi/10.1515/9781400841035/html.

49. Bretó C, Ionides EL. 2011 Compound Markov counting processes and their applications to modeling infinitesimally over-dispersed systems. *Stoch. Process. Their Appl.* **121**, 2571–2591. (doi:10.1016/j.spa.2011.07.005)

50. Ionides EL, Ning N, Wheeler J. In press. An Iterated block particle filter for inference on coupled dynamic systems with shared and unit-specific parameters. *Stat. Sin.* (doi:10.5705/ss.202022.0188)

51. Ning N, Ionides EL. 2023 Iterated block particle filter for high-dimensional parameter learning: beating the curse of Dimensionality. *J. Mach. Learn. Res.* **24**, 1–76. https://www.jmlr.org/papers/volume24/21-1253/21-1253.pdf

52. Brauner JM *et al*. 2021 Inferring the effectiveness of government interventions against COVID-19. *Science* **371**, eabd9338. (doi:10.1126/science.abd9338)

53. Park M, Cook AR, Lim JT, Sun Y, Dickens BL. 2020 A systematic review of COVID-19 epidemiology based on current evidence. *J. Clin. Med.* **9**, 967. (doi:10.3390/jcm9040967)

54. Read JM, Bridgen JR, Cummings DA, Ho A, Jewell CP. 2021 Novel coronavirus 2019-nCoV (COVID-19): early estimation of epidemiological parameters and epidemic size estimates. *Phil. Trans. R. Soc. B* **376**, 20200265. (doi:10.1098/rstb.2020.0265)

55. Subramanian R, He Q, Pascual M. 2021 Quantifying asymptomatic infection and transmission of COVID-19 in New York City using observed cases, serology, and testing capacity. *Proc. Natl Acad. Sci. USA* **118**, e2019716118. (doi:10.1073/pnas.2019716118)

56. Ioannidis JP, Cripps S, Tanner MA. 2022 Forecasting for COVID-19 has failed. *Int. J. Forecast.* **38**, 423–438. (doi:10.1016/j.ijforecast.2020.08.004)

57. Yu B, Kumbier K. 2020 Veridical data science. *Proc. Natl Acad. Sci. USA* **117**, 3920–3929. (doi:10.1073/pnas.1901326117)

58. Gneiting T, Raftery AE. 2007 Strictly proper scoring rules, prediction, and estimation. *J. Am. Stat. Assoc.* **102**, 359–378. (doi:10.1198/016214506000001437)

59. Faraway JJ. 2014 *Linear models with R*. Boca Raton, FL: CRC Press.

60. Gentleman R, Temple Lang D. 2007 Statistical analyses and reproducible research. *J. Comput. Graph. Stat.* **16**, 1–23. (doi:10.1198/106186007X178663)

61. Ionides EL, Breto C, Park J, Smith RA, King AA. 2017 Monte Carlo profile confidence intervals for dynamic systems. *J. R. Soc. Interface* **14**, 1–10. (doi:10.1098/rsif.2017.0126)

62. Ning N, Ionides EL, Ritov Y. 2021 Scalable Monte Carlo inference and rescaled local asymptotic normality. *Bernoulli* **27**, 2532–2555. (doi:10.3150/20-BEJ1321)

63. R Core Team. 2021 *R: a language and environment for statistical computing*. Vienna, Austria: R foundation for statistical computing.

64. Asfaw K, Park J, King AA, Ionides EL. 2024 A tutorial on spatiotemporal partially observed Markov process models via the R package spatPomp. *arXiv*. 2101.01157. (doi:10.48550/arXiv.2101.01157)

65. Jifan L. 2021 Metapoppkg. *GitHub*. See https://github.com/jifanli/metapoppkg.

66. Jifan L, Edward I. 2023 Jifanli/metapoppkg: Arxiv submission. *Zenodo*. (doi:10.5281/zenodo.10149232)

67. Jifan L. 2024 Metapop article. *GitHub*. See https://github.com/jifanli/metapop_article.

68. Jifan L, Edward I. 2023 Jifanli/metapop_article: v1.0.1. *Zenodo*. (doi:10.5281/zenodo.10149258)

69. Li J, Ionides E, King AA, Pascual M, Ning N. 2024 Supplementary material from: Inference on Spatiotemporal Dynamics for coupled biological populations. Figshare. (doi:/10.6084/m9.figshare.c.7302773)