Statistica Sinica Preprint No: SS-2022-0188	
Title	An Iterated Block Particle Filter for Inference on Coupled Dynamic Systems With Shared and Unit-Specific Parameters
Manuscript ID	SS-2022-0188
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202022.0188
Complete List of Authors	Edward Ionides, Ning Ning and Jesse Wheeler
Corresponding Authors	Edward Ionides
E-mails	ionides@umich.edu

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

Statistica Sinica

An iterated block particle filter for inference on coupled dynamic systems with shared and unit-specific parameters

Edward L. Ionides[1], Ning Ning[2] and Jesse Wheeler[1]

University of Michigan, Department of Statistics[1]

Texas A&M University, Department of Statistics[2]

For *Statistica Sinica* special edition on Recent developments of sequential Monte Carlo methods and their applications

Abstract: We consider inference for a collection of partially observed stochastic interacting nonlinear dynamic processes. Each process is identified with a label, called its unit. Here, our primary motivation arises in biological metapopulation systems, in which a unit corresponds to a spatially distinct sub-population. Metapopulation systems are characterized by strong dependence over time within a single unit, and relatively weak interactions between units. These properties make block particle filters effective for simulation-based likelihood evaluation. Iterated filtering algorithms can facilitate likelihood maximization for simulation-based filters. Here, we introduce an iterated block particle filter that can be applied when parameters are unit-specific or shared between units. We demonstrate the proposed algorithm by performing inference on a coupled epidemiological model describing spatiotemporal measles case report data for 20 towns.

Key words and phrases: sequential Monte Carlo; metapopulation; spatiotemporal; maximum likelihood estimation; partially observed Markov process.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

1. Introduction

Statistical inference for high-dimensional partially observed nonlinear dynamic systems arises in various scientific contexts. Massive models and data sets are considered in the geophysical sciences, carried out under the name of data assimilation (Evensen, 2009). Population models in ecology and epidemiology can be characterized by high levels of stochasticity, nonlinearity, measurement error, and model uncertainty, leading to challenges of a somewhat different nature to those of geophysical models. In addition, biological population systems may have a low population count, owing to a local introduction or fade-out of one or more constituent species. Such situations may require models with integer-valued counts, rather than continuous population approximations. Collections of biological populations measured at different spatial locations may have spatial interactions in addition to local population dynamics; such collections are called a metapopulation. The study of spatiotemporal disease dynamics has motivated research into inference for metapopulation systems (Xia et al., 2004; Li et al., 2020; Park and Ionides, 2020; Ionides et al., 2021; Cauchemez et al., 2008; Bjørnstad et al., 2019).

Until recently, statistical inference for partially observed nonlinear biological systems was an open methodological challenge, even in the time

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

series case (Bjørnstad and Grenfell, 2001). Advances in Monte Carlo methods based on particle filters have made inference accessible in many low-dimensional problems (Doucet and Johansen, 2011; Kantas et al., 2015; King et al., 2016). However, the curse of dimensionality (Bengtsson et al., 2008) prevents application of the basic particle filter on metapopulations with more than a few units. Methods based on improving the proposal distribution for the particle filter may not fully resolve the curse of dimensionality (Snyder et al., 2015). Previous Monte Carlo methods which can provably beat the curse of dimensionality may have limited applicability (Beskos et al., 2017; Park and Ionides, 2020; Ionides et al., 2021). Consequently, state-of-the-art scientific analysis for metapopulation dynamics depends on ensemble Kalman filter (EnKF) methods (Li et al., 2020). EnKF algorithms scale well, but are founded on an approximation that can be unsuitable for discrete populations with fade-out and re-introduction dynamics and other highly non-Gaussian features (Ionides et al., 2021).

In Section 2, we propose an algorithm for inference on metapopulation dynamics, which we call an iterated block particle filter (IBPF). IBPF algorithms combine an iterated filtering likelihood maximization technique (Ionides et al., 2015) with a block particle filter (BPF) (Rebeschini and Van Handel, 2015). Iterated filtering algorithms use parameter perturba-

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

tions to coerce a filtering algorithm into exploring the parameter space. BPF algorithms address the curse of dimensionality by modifying the resampling step of a particle filter to resample independently on blocks that form a partition of the collection of units. A previous IBPF algorithm was developed by Ning and Ionides (2023) for the particular case in which all estimated parameters are unit-specific, that is, the dynamics and measurement process for a unit u are determined by a vector of parameters ψ_u specific to unit u. In Section 2, we provide a formal meaning of this assertion, together with the pseudocode for our algorithm. We propose an extension of the aforementioned IBPF which additionally allows us to estimate a vector of shared parameters, ϕ. In this case, the full parameter vector is $\theta = (\phi, \psi_{1:U})$, where the U units are named $\{1, \ldots, U\}$, which we denote by $1:U$. Ning and Ionides (2023) develop a theoretical justification for their algorithm; however, our new IBPF currently relies on empirical support only.

There may be scientific interest in which parameters in a metapopulation system are best understood as being unit-specific, and which can reasonably be modeled as being shared between units. Another relevant possibility is that a parameter may differ between units as a shared function of unit-specific covariates. Formally, this is a special case of a shared

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

parameter. Addressing these issues is also a prerequisite for studying questions about the coupling of metapopulation systems using model-based inference from spatiotemporal data. In Section 3.3, we will show empirically that our IBPF algorithm is applies to an inference challenge in epidemiological metapopulation dynamics. Our demonstration focuses on a data set of weekly measles incidence in 20 towns in the United Kingdom (UK) during the pre-vaccination era (He et al., 2010), modeled using a previously studied metapopulation model (Park and Ionides, 2020; Ionides et al., 2021). Measles case reports are a longstanding benchmark problem for inference on biological dynamics, motivating the development of time series methodologies and, more recently, the progression from single populations to metapopulation systems. Unlike previous attempts on sequential Monte Carlo inference for metapopulation models, we show that our algorithm can provide practical plug-and-play, likelihood-based inference when the parameters are either shared between units or differ between units. Demonstrating a solution to this open problem provides numerical evidence substituting for numerical comparisons with alternative methods.

Our data analysis results do not fully resolve open questions about what models for coupling between towns are supported by the data, and which parameters should be modeled as unit-specific. Rather, we demonstrate

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

steps toward this research goal. We use a simulation study, discussed in Section 3.2, to show that our methodology can deliver a good approximation to the maximum likelihood estimate (MLE) when fitting the model used to simulate the data. This allows us to interpret our data analysis results as evidence of model misspecification, providing a guide for future investigations of these data, as well as tools to carry out such investigations.

Optimization of high-dimensional, non-convex and potentially multimodal functions, evaluated using stochastic methods, is not straightforward, even when it is possible to evaluate the function within an acceptable level of error. Therefore, we discuss approaches that assist noisy likelihood searches, and suggest diagnostic plots to assess their success.

2. An IBPF algorithm for likelihood maximization

A latent Markov process is denoted by $\{\boldsymbol{X}_n, n \in 0:N\}$, with $\boldsymbol{X}_n = X_{1:U,n}$ taking values in a product space $\mathbb{X}^U$. We define set-valued subscripts by $X_A = \{X_a, a \in A\}$ and $X_{A,B} = \{X_{a,b}, a \in A, b \in B\}$. The discrete time process $\boldsymbol{X}_{0:N}$ may arise from a continuous-time Markov process $\{\boldsymbol{X}(t), t_0 \le t \le t_N\}$ observed at times $t_{1:N}$, in which case, we set $\boldsymbol{X}_n = \boldsymbol{X}(t_n)$. The initial value $\boldsymbol{X}_0$ may be stochastic or deterministic. Observations are made on each unit, modeled by an observable process

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

$\boldsymbol{Y}_{1:N} = Y_{1:U,1:N}$ that takes values at each time n in a product space $\mathbb{Y}^U$. Observations are modeled as conditionally independent, given the latent process. The conditional independence of measurements applies over both time and the unit structure; thus, the collection $\{Y_{u,n}, u \in 1:U, n \in 1:N\}$ is conditionally independent, given $\{X_{u,n}, u \in 1:U, n \in 1:N\}$. We suppose the existence of a joint density $f_{\boldsymbol{X}_{0:N},\boldsymbol{Y}_{1:N}}$ for $X_{1:U,0:N}$ and $Y_{1:U,1:N}$ with respect to some appropriate measure, following a notational convention that the subscripts of f denote the joint or conditional density under consideration. We suppose that f depends on a real-valued parameter vector $\theta = (\phi, \psi_{1:U})$, which we write as $\theta = \theta_{1:D}$ when we wish to concatenate the shared and unit-specific parameters into a single vector of length D. The data are $y^*_{u,n}$, for unit u at time n. This model is a special case of a partially observed Markov process (POMP), also known as a state-space model or hidden Markov model. The additional unit structure, not generally required for a POMP, is appropriate for modeling interactions between units characterized by a spatial location; thus, we call the model a Spat-POMP. For metapopulation models, the units are not, in general, arranged on a spatial grid, but instead comprise a collection of spatially distributed population centers.

A numerical challenge of fundamental statistical relevance is maxi-

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

mizing the log-likelihood function of the data, given the model, $\ell(\theta) = \log f_{\boldsymbol{Y}_{1:N}}(\boldsymbol{y}^*_{1:N};\theta)$. Numerical evaluation of the likelihood function is closely related to the filtering problem of evaluating $f_{\boldsymbol{X}_n|\boldsymbol{Y}_{1:n}}(\boldsymbol{x}_n \,|\, \boldsymbol{y}^*_{1:n};\theta)$. If the dynamic model is extended to include the parameters as latent variables, the filtering problem leads to the Bayesian posterior distribution, though regularization is required to make the calculation numerically tractable using Monte Carlo methods (Kitagawa, 1998; Liu and West, 2001). Iterating this Bayesian calculation recursively targets an MLE, a strategy known as data cloning (Lele et al., 2007, 2010). Adding noise to perturb the parameters in the extended model at each time point can stabilize the numerics, while still being able to approximate the MLE (Ionides et al., 2006, 2011, 2015). Many variations on this idea have been developed using different filter methods (Park and Ionides, 2020; Li et al., 2020; Ionides et al., 2021; Manoli et al., 2015) or employing different perturbation systems (Doucet et al., 2013; Nguyen and Ionides, 2015).

A direct approach to iterating a BPF for parameter estimation is to resample the extended model independently on each block, yielding separate collections of parameters for each block. Ning and Ionides (2023) prove that this IBPF algorithm targets the MLE for the special case where each parameter is localized to an individual unit, that is, when all parameters are

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

unit-specific. Formally, we say that a parameter for a discrete-time SpatPOMP is specific to a unit u if it is involved in specifying the measurement density $f_{Y_{u,n}|X_{u,n}}$ or the transition density $f_{X_{u,n+1}|\boldsymbol{X}_n}$, for some n, and it is not involved in $f_{Y_{v,m}|X_{v,m}}$ or $f_{X_{v,m+1}|\boldsymbol{X}_m}$, for any $v \neq u$ and any m. For a continuous-time SpatPOMP, we replace the requirement on $f_{X_{u,n+1}|\boldsymbol{X}_n}$ with an equivalent requirement on a numerical solution over a small time increment δ, as in equation (2.1). A parameter that is not unit-specific is said to be shared between units. There are intermediate possibilities, where a parameter is shared for only a subset of all units, but such a parameter is formally classified as shared. The special case in which all parameters are unit-specific may occur, but models typically have some shared parameters that arise in transition densities and/or measurement densities for multiple units.

In our approach to iterated filtering for shared parameters, we construct an extended model, within which, the shared parameters are represented as a unit-specific parameter that happens to be constant across units. We construct a spatiotemporally extended model by supposing that a numerical solution for the transition from time t to time $t + \delta$ for each unit u has the

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

functional form

$$X_u(t+\delta) = X_u(t) + Q_u\big(\boldsymbol{X}(t), \boldsymbol{\eta}_t, \phi, \psi_u, t, \delta\big), \tag{2.1}$$

where the random vector $\boldsymbol{\eta}_t = \eta_{1:U,t}$ is shared, for all $u \in 1:U$, and does not depend on θ. If a representation (2.1) exists, an extended model is defined by replacing ϕ with $\phi_u(t)$ and ψ_u with $\psi_u(t)$. Equation (2.1) implicitly defines a continuous-time extended model by the limit as $\delta \to 0$, when that limit exists, but for simulation-based methods, a numerical solution is of more immediate concern than this limit. Admitting a minor abuse of notation, we subsequently use the density f to denote both the original model and its extension for spatiotemporally varying parameters, with the context determining which one is intended.

In some situations, the extended model may be problematic; for example, it could break conservation laws obeyed by the original system. In other situations, the extended model may make scientific sense in its own right; for example, in biological metapopulation systems, it might be scientifically meaningful to consider a model in which there is variation over space and time in the parameters that describe the local dynamics. Here, we focus on the hypothesis that some parameters are fixed across space and

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

time, but the specification in (2.1) requires that this hypothesis be nested within a more flexible alternative.

Statistica Sinica

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

Algorithm 1: IBPF.

Inputs: simulator for the extended model, $f_{\boldsymbol{X}_n|\boldsymbol{X}_{n-1}}(\boldsymbol{x}_n \,|\, \boldsymbol{x}_{n-1}\,;\boldsymbol{\theta})$, and initialization, $f_{\boldsymbol{X}_0}(\boldsymbol{x}_0\,;\boldsymbol{\theta})$; evaluator for $f_{Y_{u,n}|X_{u,n}}(y_{u,n} \,|\, x_{u,n}\,;\theta)$; data, $\boldsymbol{y}^*_{1:N}$; number of particles, J; blocks, $\mathcal{B}_{1:K}$; initial parameter swarm with decomposition into shared and unit-specific parameters, $\Theta^{0,j}_u = (\Phi^{0,j}_u, \Psi^{0,j}_u)$; random walk perturbation, $\sigma_{d,n}$; cooling rate, a; number of iterations, M; spatial autoregression, r.

for m in $1:M$ **do**
  Perturb parameters: $\Theta^{F,m,j}_{u,0} \sim \mathcal{N}\left(\Theta^{m-1,j}_u\,; \sigma_0^2 a^{2m/50}\right)$
  Initialization: simulate $\boldsymbol{X}^{F,j}_0 \sim f_{\boldsymbol{X}_0}\left(\cdot\,; \Theta^{F,m,j}_{1:U,0}\right)$
  for n in $1:N$ **do**
    Perturb parameters: $\Theta^{P,m,j}_{u,n} \sim \mathcal{N}\left(\Theta^{F,m,j}_{u,n-1}, \sigma_n^2 a^{2m/50}\right)$
    Prediction simulation: $\boldsymbol{X}^{P,j}_n \sim f_{\boldsymbol{X}_n|\boldsymbol{X}_{n-1}}\left(\cdot\,|\boldsymbol{X}^{F,j}_{n-1}; \Theta^{P,m,j}_{1:U,n}\right)$
    for k in $1:K$ **do**
      Block prediction weights:
      $$w^P_{n,j,k} = \prod_{u\in\mathcal{B}_k} f_{Y_{u,n}|X_{u,n}}\left(y^*_{u,n} \,|\, X^{P,j}_{u,n}\,; \Theta^{P,m,j}_{u,n}\right)$$
      Normalize weights: $\tilde{w}_{n,j,k} = w^P_{n,j,k} \Big/ \sum_{i=1}^J w^P_{n,i,k}$
      Select resample indices: $i_{1:J,k}$ with $\mathbb{P}\left[i_{j,k} = k\right] = \tilde{w}_{n,j,k}$
      $X^{F,j}_{\mathcal{B}_k,n} = X^{P,i_{j,k}}_{\mathcal{B}_k,n}$, $\Theta^{F,j}_{\mathcal{B}_k,n} = \Theta^{P,i_{j,k}}_{\mathcal{B}_k,n} = \left(\Phi^{P,i_{j,k}}_{\mathcal{B}_k,n}, \Psi^{P,i_{j,k}}_{\mathcal{B}_k,n}\right)$
      block mean of shared parameters: $\mu_{k,n} = J^{-1}\sum_{j=1}^J \Phi^{F,j}_{\mathcal{B}_k,n}$
    end
    overall mean of shared parameters: $\mu_n = K^{-1}\sum_{k=1}^K \mu_{k,n}$
    autoregressive correction: $\Phi^{F,j}_{\mathcal{B}_k,n} = \Phi^{F,j}_{\mathcal{B}_k,n} + r\left(\mu_n - \mu_{k,n}\right)$
  end
  $\Theta^{m,j}_u = \Theta^{F,m,j}_{u,N}$
end

output: IBPF parameter swarm, $\Theta^{M,j}_u$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

The IBPF algorithm described in Algorithm 1 carries out a BPF on this extended model. The extended parameters are given independent perturbations, but a spatial autoregressive step pulls the values of the shared parameters toward their mean over the units. Indeed, this autoregressive step is the only difference between Algorithm 1 and the algorithm for unit-specific parameters proposed by Ning and Ionides (2023). This algorithm, in turn, is essentially the IF2 algorithm of Ionides et al. (2015), with the particle filter replaced with the BPF of Rebeschini and Van Handel (2015).

Algorithm 1 assumes implicit loops over j in $1:J$ and u in $1:U$, $\mathcal{N}(\mu, \Sigma)$ denotes the normal distribution with mean μ and variance matrix Σ, and σ_n is a $D \times D$ diagonal matrix with entries $\sigma_{d,n}$. The blocks $\mathcal{B}_{1:K}$ are a partition of $1:U$.

The pseudocode in Algorithm 1 represents our implementation of an IBPF as the R function `ibpf`, which we have contributed to the open-source package `spatPomp` (Asfaw et al., 2021, 2023). Additionally, the source code for all results presented here is available at `https://github.com/ionides/ibpf_article`. Various generalizations of this implementation are possible. For example, iterated filtering theory does not rely on parameter perturbations following the normal distribution (Ionides et al., 2015), though, in practice, we transform the parameters to facilitate this convenient choice

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

(see Section S2).

2.1 Algorithmic parameters

The model parameters are optimized on a transformed scale for which unit variation is scientifically meaningful. In practice, this means working with positive parameters on a log scale, and with $(0, 1)$ interval-valued parameters on a logistic scale. We follow standard iterated filtering practice by using independent random walks for each parameter on this transformed scale (King et al., 2016). We find that the large number of parameters following a random walk, in the presence of unit-specific parameters, can require considerably smaller random walk standard deviations than the values around $\sigma_{d,n} = 0.02$ (i.e., 2% perturbation per time point) that have been employed for iterated filtering of time series models. After experimentation, we used $\sigma_{d,n} = 0.005$ for the initial search, and $\sigma_{d,n} = 0.00125$ for subsequent refinements. Occasionally, a parameter that can be estimated precisely from the data can benefit substantially from a smaller perturbation. This is the case for one parameter in our measles analysis, an exponent α, for which the scale of uncertainty is an order of magnitude smaller than that of the other parameters; therefore, we scaled $\sigma_{d,n}$ for this parameter accordingly. In principle, $\sigma_{d,n}$ can be a function of n and the parameter,

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

d. The most common reason for using this flexibility is to avoid perturbing parameters during time intervals in which we have no information about these parameters. Following an evolutionary analogy, evolution cannot operate effectively if there is mutation, but no selection. For an initial value parameter, that specifies only the latent process value at the initial time t_0, we use $\sigma_{d,n} = 0$ for $n \geq 1$, and we double the value of $\sigma_{d,n}$, for $n = 0$.

We use $J = 4000$ particles, and set the cooling rate parameter to $a = 0.5$, corresponding to a 1% reduction in the random walk standard deviation at each iteration. We set $M = 100$ optimization iterations, chosen as an empirically assessed compromise between the effort spent on each search and the number of searches conducted. An additional discussion of algorithmic parameters is given in Section S2. The only additional algorithmic parameter over previous iterated filtering algorithms is the spatial autoregressive parameter, r. The results of our numerical experiments suggest that the performance is not sensitive to the choice of $r > 0$ (see the Supplementary Material, Figure S1). Consequently, here we use $r = 0.1$.

Applying an IBPF to real data forces us to address problems related to model development, model misspecification, and performance in the presence of outliers. Before doing so, we use simulated data to demonstrate the capabilities of the methodology on a correctly specified model.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

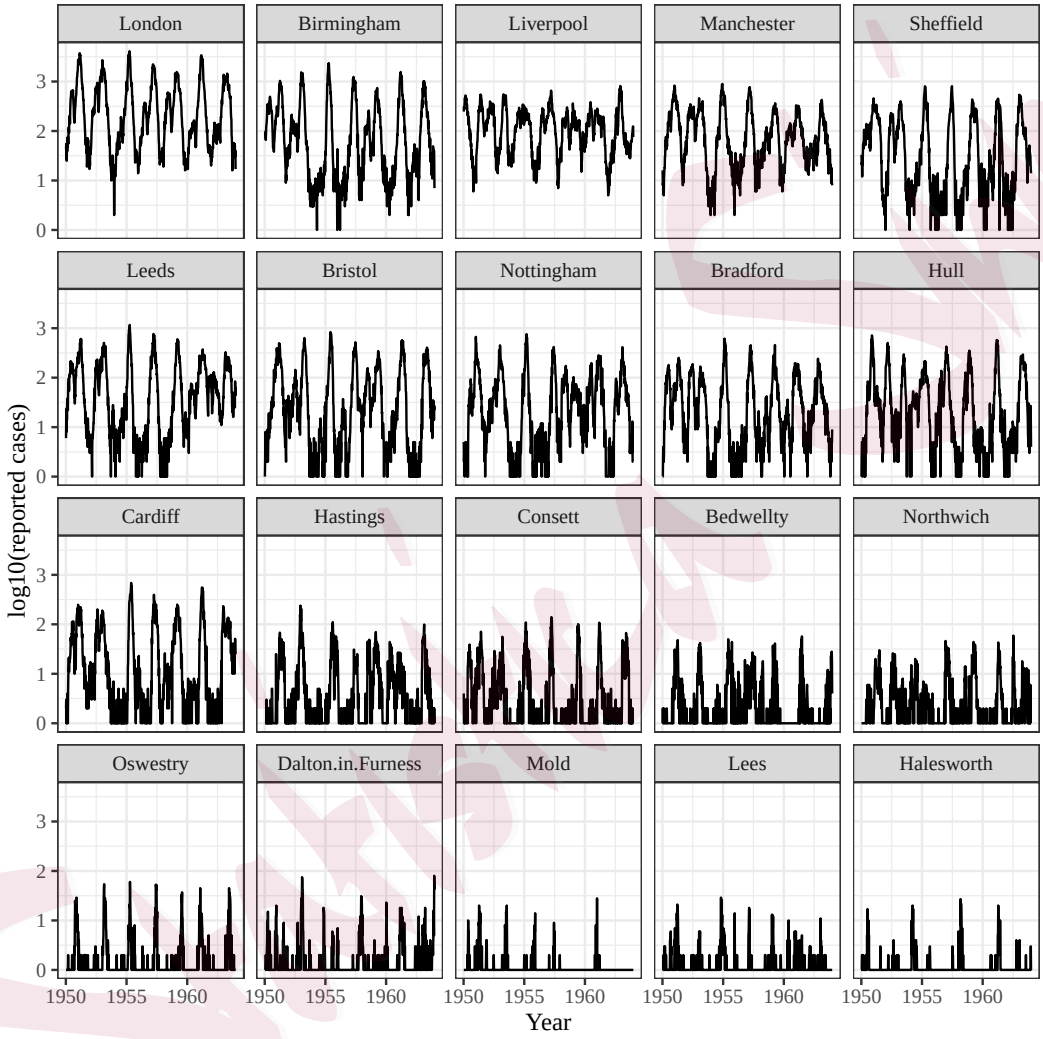

Figure 1: Weekly measles case reports for 20 UK towns.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

3. Testing an IBPF on a measles transmission model

Measles transmission is a useful example of epidemiological dynamics (and therefore also ecological dynamics, for a host-pathogen ecosystem), with plentiful case report data and relatively simple biology. Model-based analyses of measles time series data have led to a better understanding of the seasonality of infectious diseases (Fine and Clarkson, 1982), critical community size (Bartlett, 1957), the recognition that relatively simple mechanistic models can provide a remarkably good description of the dynamics (Earn et al., 2000), and other foundational research on disease dynamics. Some progress has been made on building and fitting spatiotemporal models for measles; see, for example, Xia et al. (2004), Eggo et al. (2011), Bjørnstad et al. (2019), and Becker et al. (2020). However, the lack of suitable methodology to fit and assess a flexible class of coupled models is an obstacle to further progress (Becker et al., 2020). Previous methodological research has used spatiotemporal measles models as a test problem (Park and Ionides, 2020; Ionides et al., 2021). However, these methods have fallen short as tools for data analysis, owing to numerical considerations. Bearing all this in mind, measles provides a natural testing ground for our new methodologies. We demonstrate that we now have the tools to carry out likelihood maximization (and therefore, in principle, profile likelihood confidence in-

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

terval construction and likelihood-based model selection) on mechanistic statistical models that are appropriate for spatiotemporal metapopulation disease data.

3.1 The measles data, a model, and three submodels

We set ourselves the task of fitting a spatiotemporal model to the case reports for 20 towns studied by He et al. (2010). We need to be able to handle discrete case counts, which vary from zero to thousands of cases per week. We consider a model with the same structure as that of He et al. (2010), namely, a Markov chain with gamma noise on the infection rate, but with an additional term for the transmission between cities. This requirement limits us to plug-and-play methodologies, which are those that require simulation from the latent process model, but not the ability to evaluate transition densities (Bretó et al., 2009; He et al., 2010).

Some previous analyses have used counts aggregated over two-week intervals (Park and Ionides, 2020; Ionides et al., 2021), because these were available for more cities, but our goal here is to extend the analysis of He et al. (2010). Apart from this, our model matches that of Ionides et al. (2021) and, for completeness, we repeat the description here. We compartmentalize the population of each town into susceptible (S), exposed

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

(E), infectious (I), and recovered/removed (R) categories. The numbers of individuals in each compartment for town u at time t are denoted by integer-valued random variables $S_u(t)$, $E_u(t)$, $I_u(t)$, and $R_u(t)$, respectively. The population dynamics are written in terms of the counting processes $N_{Q_1Q_2,u}(t)$ enumerating the cumulative transitions in town u, up to time t, from compartment Q_1 to Q_2. Here, $Q_1, Q_2 \in \{S, E, I, R, B, D\}$, with B denoting a source compartment for immigration or birth, and D denoting a sink compartment for emigration or death. We enumerate the $U = 20$ towns studied by He et al. (2010) in decreasing size, so that $u = 1$ corresponds to London. Our model is described by the following system of stochastic differential equations, for $u = 1, \ldots, U$:

$$\begin{aligned} dS_u(t) &= dN_{BS,u}(t) - dN_{SE,u}(t) - dN_{SD,u}(t), \\ dE_u(t) &= dN_{SE,u}(t) - dN_{EI,u}(t) - dN_{ED,u}(t), \\ dI_u(t) &= dN_{EI,u}(t) - dN_{IR,u}(t) - dN_{ID,u}(t). \end{aligned}$$

The total population $P_u(t) = S_u(t) + E_u(t) + I_u(t) + R_u(t)$ is calculated by smoothing census data, and is treated as known. The number of recovered individuals $R_u(t)$ in town u is therefore defined implicitly. Furthermore, $N_{SE,u}(t)$ is modeled as a negative binomial death process (Bretó et al.,

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

2009; Bretó and Ionides, 2011), defined by a rate, $\mu_{SE,u}(t)$, specified as

$$\mu_{SE,u} = \beta_u \left[\left(\frac{I_u + \iota_u}{P_u} \right)^{\alpha_u} + \sum_{\tilde{u} \neq u} \frac{v_{u\tilde{u}}}{P_u} \left\{ \left(\frac{I_{\tilde{u}}}{P_{\tilde{u}}} \right)^{\alpha_u} - \left(\frac{I_u}{P_u} \right)^{\alpha_u} \right\} \right] \frac{d\Gamma_{SE,u}}{dt}.$$

Here, the time dependence of $\mu_{SE,u}$, β_u, I_u, and P_u is suppressed; α_u is a mixing exponent modeling inhomogeneous contact rates within a town; ι_u models immigration of infected individuals; $d\Gamma_{SE,u}/dt$ is gamma white noise, with intensity parameter $\sigma_{SE.u}$; and β_u models the seasonal transmission driven by high contact rates between children at school,

$$\beta_u(t) = \begin{cases} \left(1 + h_u(1-s)s^{-1}\right) \overline{\beta}_u & \text{during school term,} \\ \left(1 - h_u\right) \overline{\beta}_u & \text{during vacation,} \end{cases}$$

where $s = 0.759$ is the proportion of the year taken up by the school terms, $\overline{\beta}_u$ is the mean transmission rate, and h_u measures the reduction in transmission during school holidays.

The number of travelers from town u to $\tilde{u}$ is denoted by $v_{u\tilde{u}}$. Here, $v_{u\tilde{u}}$ is constructed using the gravity movement model of Xia et al. (2004), given by

$$v_{u\tilde{u}} = G_u \cdot \frac{\overline{d}}{\overline{\text{pop}}^2} \cdot \frac{\text{pop}_u \cdot \text{pop}_{\tilde{u}}}{d_{u\tilde{u}}},$$

where G_u is the *gravitational constant*, $d_{u\tilde{u}}$ is the distance between town u

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

and town $\tilde{u}$, pop_u is the average of $P_u(t)$ over time, $\overline{\text{pop}}$ is the average of pop_u across towns, and $\overline{d}$ is the average of $d_{u\tilde{u}}$ across all pairs of towns. The transition processes $N_{EI,u}(t)$, $N_{IR,u}(t)$, and $N_{QD,u}(t)$, for $Q \in \{S, E, I, R\}$, are modeled as conditional Poisson processes with per-capita rates $\mu_{EI,u}$, $\mu_{IR,u}$, and $\mu_{QD,u}$, respectively, and we fix $\mu_{QD,u} = (50 \text{ year})^{-1}$. A fraction c of births enter the susceptible cohort on the school admission day, and hence the birth process $N_{BS,u}(t)$ is an inhomogeneous Poisson process with rate $\mu_{BS,u}(t-\Delta)\big[(1-c)+c\delta(t-t_a)\big]$, where $\mu_{BS,u}(t)$ is specified by interpolated census data, $t_a = \lfloor t \rfloor + 252/365$ is the admission date for the year containing t, δ is a Dirac delta function, and $\Delta = 4$ year is a fixed delay between birth and entry into a high-transmission school community.

To describe the measurement process, let $Z_{u,n} = N_{IR,u}(t_n) - N_{IR,u}(t_{n-1})$ be the number of removed infected individuals in the nth reporting interval. Suppose that cases are quarantined once they are identified, so that reported cases comprise a fraction ρ of these removal events. The case report $y^*_{u,n}$ is modeled as a realization of a discretized conditionally Gaussian random variable $Y_{u,n}$, defined for $y > 0$ via

$$\begin{aligned}\mathbb{P}\big[Y_{u,n}{=}y \mid Z_{u,n}{=}z\big] \;&=\; \Phi\big(y+0.5; \rho_u z, \rho_u(1-\rho_u)z + \tau_u^2\rho_u^2 z^2\big) \\ &\quad -\Phi\big(y-0.5; \rho_u z, \rho_u(1-\rho_u)z + \tau_u^2\rho^2 z^2\big), \qquad (3.1)\end{aligned}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

where $\Phi(\cdot;\mu,\sigma^2)$ is the $\mathcal{N}(\mu,\sigma^2)$ cumulative distribution function, and τ models overdispersion relative to the binomial distribution. For $y=0$, we replace $y-0.5$ with $-\infty$ in (3.1). Three data points are treated as missing by He et al. (2010), owing to presumed recording errors. We follow this decision, implemented by including a special value NA in $\mathbb{Y}$, and setting $Y_{u,n}$ to NA with probability one when $y^*_{u,n}$ is missing.

We have written the model with all parameters unit-specific, so $\theta=\psi_{1:U}$ with $\psi_u = \big(\overline{\beta}_u, \mu_{EI,u}, \mu_{IR,u}, \rho_u, \tau_u, \sigma_{SE,u}, G_u, \iota_u, h_u, \alpha_u, S_{0,u}, E_{0,u}, I_{0,u}\big)$, for $u \in 1:U$. This defines an extended model for implementing an IBPF with shared parameters, as in (2.1). It also serves another function by allowing us address a key data analysis question of which parameters should be shared between units and which should be unit-specific.

The data are shown in Fig 1, and simulations from the model are shown in Fig 2. Parameters for the simulated model were based on the analysis of individual towns by He et al. (2010). In order to investigate estimation of either shared or unit-specific parameters, we conduct the simulation with all parameters being shared. To find a shared parameter vector capable of providing a reasonable representation of all the towns simultaneously, we experimented to look for a visual match between the data and the simulations. The parameters used are reported in the Supplementary Material

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

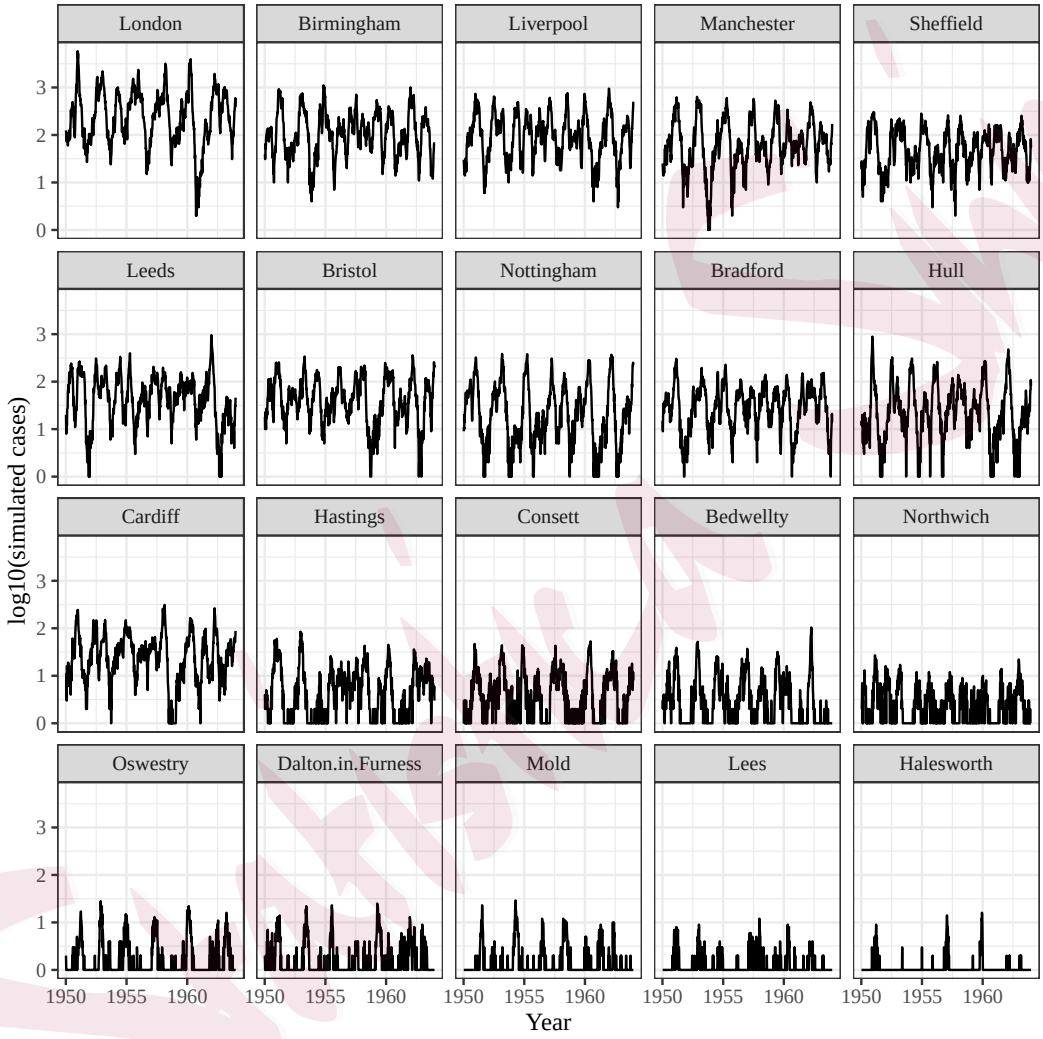

Figure 2: Simulated weekly measles case reports.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

(Section S1).

Estimated parameter values may have scientific interest, but we focus on the statistical task of likelihood maximization. In the presence of weak identifiability, small differences in likelihood could lead to large differences in parameter estimates. In such situations, a scientist may choose to investigate what functions of the parameters can be inferred accurately without adding additional assumptions. Alternatively, they may choose to investigate the consequences of placing constraints on some parameters in order to improve the identifiability of the remainder. Likelihood maximization permits such investigations, but is beyond the scope of this study. In the Supplementary Material, we provide the parameter values used for the simulation study and those obtained by the likelihood maximization.

To facilitate data analysis, our methodology needs to operate across the full spectrum of decisions on shared versus unit-specific parameter designations. We therefore test our method on three submodels: $\mathcal{A}$ has mostly shared parameters, with only the initial values and the reporting rate unit-specific; $\mathcal{B}$ has every parameter unit-specific; and $\mathcal{C}$ has every parameter unit-specific, and the dynamic coupling (in our context, the movement of infected individuals between towns) is replaced with an external forcing of each unit (in our context, the immigration rate of infected individuals from

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

outside the study population). Section S1 defines the submodels in further detail. Submodel $\mathcal{C}$ provides a useful point of reference, because it is a special case of a PanelPOMP model (Bretó et al., 2019), and also can be analyzed as a collection of separate POMP models. We set up our model so that $\mathcal{C}$ matches the analysis of He et al. (2010). One may expect that methods that take advantage of the special structure of $\mathcal{C}$ should outperform more general methods that permit coupling between units. Thus, we expect the SpatPOMP methods to be less efficient numerically than application of POMP methods separately to each unit. Here we answer the following questions: How much less efficient are the SpatPOMP methods? Are simulation-based SpatPOMP inference methods practical for situations such as the measles model of He et al. (2010)?

The simulated model is drawn from model $\mathcal{A}$, which is nested within model $\mathcal{B}$, but not within $\mathcal{C}$. For models $\mathcal{B}$ and $\mathcal{C}$, in which all parameters are unit-specific, we must estimate $20 \times 13 = 260$ parameters. This greatly exceeds the seven shared parameters fitted by Ionides et al. (2021) for a measles metapopulation model, which required significant computational effort. An iterated guided intermediate resampling filter (IGIRF) algorithm was used to fit nine shared parameters and three unit-specific initial value parameters in a measles metapopulation model (Park and Ionides, 2020).

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

The latter analysis uses a customized treatment for the unit-specific initial value parameters, and does not attempt to estimate other unit-specific parameters. IGIRF is sensitive to the choice of guide function, and the model-specific implementation of Park and Ionides (2020) outperforms the current generic implementation of IGIRF in `spatPomp`. Thus, to the best of our knowledge, the IBPF presented in Algorithm 1 advances the current limits on the scalability of simulation-based maximum-likelihood inference for metapopulation dynamics.

Before engaging in likelihood maximization, we first validate the likelihood evaluation; see Ionides et al. (2021) for likelihood evaluation for metapopulation models. Briefly, the basic particle filter provides a consistent evaluation of the log-likelihood in a limit with sufficient particles to make the Monte Carlo standard deviation small. This is practical only when U is small (say, $U \leq 5$), but this situation can be used to calibrate the bias induced by a BPF, which turns out to be small for our multi-town measles model when each town is its own block. It may be surprising that resampling independently on each block (which is what a BPF does) is able to capture the dependence. Heuristically, note that the dynamic dependence between blocks is maintained by a BPF, which updates particles according to the full coupled dynamics. Whether this is sufficient to obtain a good

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

approximation to the filter distribution depends on the situation, but for the specific case of metapopulation models (for which the strongest coupling is within units, rather than between units), the approximation can be empirically successful. In principle, the approximation error can be reduced by increasing the number of units in each block. However, in practice, the additional Monte Carlo variance acquired by doing so is not worthwhile when the coupling is relatively weak. The BPF log-likelihood evaluation at the true parameter value is $\ell_{\text{true}} = -40612.5$, with a Monte Carlo standard error of 0.6.

3.2 IBPF on simulated data

One of our goals is to obtain appropriate algorithmic choices for a data analysis. Thus, we seek to develop methodology that is demonstrably successful when the truth is known, before applying it to data. One could revisit the simulation study based on the data analysis in Section 3.3 using MLEs of the parameters.

In Figure 3 we investigate a sequence of successive searches for the MLE for the models $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ described previously. Each search is replicated 36 times. Search 1 was started with each parameter adjusted by a uniform $[-0.1, 0.1]$ random perturbation on an appropriate dimensionless

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

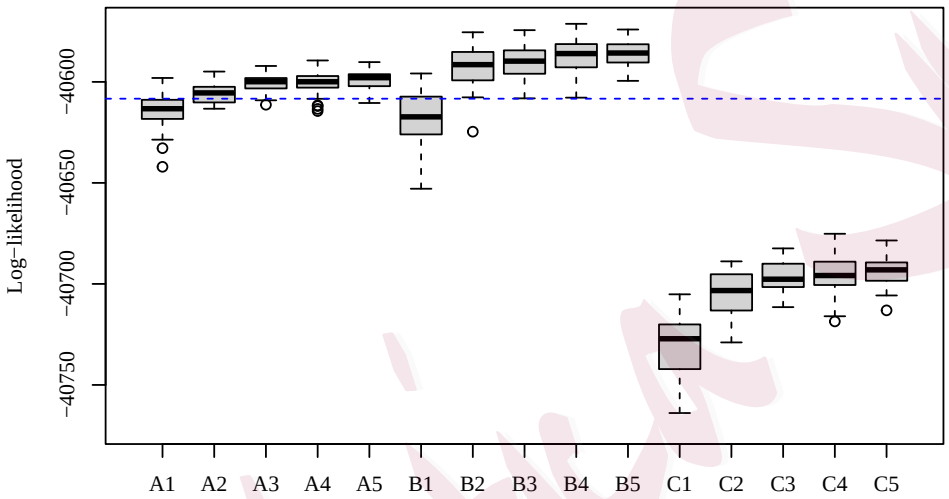

Figure 3: Fitting simulated measles data, using an initial search and four refinement steps. (A1–A5) model $\mathcal{A}$, 4×20 unit-specific parameters and nine shared parameters; (B1–B5) model $\mathcal{B}$, 13×20 unit-specific parameters and no shared parameters; (C1–C5) model $\mathcal{C}$, all unit-specific, but with immigration rather than coupling, matching He et al. (2010). The horizontal dashed line is the log-likelihood at the true parameters, evaluated using BPF.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

scale (log for nonnegative parameters, and logit for $[0, 1]$ valued parameters; see Section S2). This is a fairly small perturbation, but it is nevertheless sufficient to knock the likelihood of the parameter vector around 200 log units below the MLE (shown in Figure 4). Our goal is to show that the algorithm can succeed reliably on a relatively easy local optimization task. Subsequent searches were started with four copies of each parameter vector with an estimated likelihood in the top 25% for the previous search.

The MLE is not known exactly in this case. Wilks' theorem gives an asymptotic expectation that the log-likelihood at the MLE should be greater than the likelihood at the truth by approximately 1/2 the number of parameters, which here is $(4 \times 20 + 9)/2 = 44.5$ for $\mathcal{A}$, and $(13 \times 20)/2 = 130$ for $\mathcal{B}$. The y-axis values in Figure 3 show log-likelihoods exceeding the truth by less than this, indicating some imperfection in the Monte Carlo maximization so far as Wilks' asymptotic result holds. Despite this limitation, searches that exceed the likelihood at the truth have found inferentially plausible sets of parameters that can be used to study the likelihood surface around the maximum. For example, Monte Carlo profile confidence intervals can give proper coverage, even in the presence of considerable Monte Carlo error (Ionides et al., 2017; Ning et al., 2021).

Figure 4 shows the convergence diagnostics corresponding to $\mathcal{A}$ for the

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

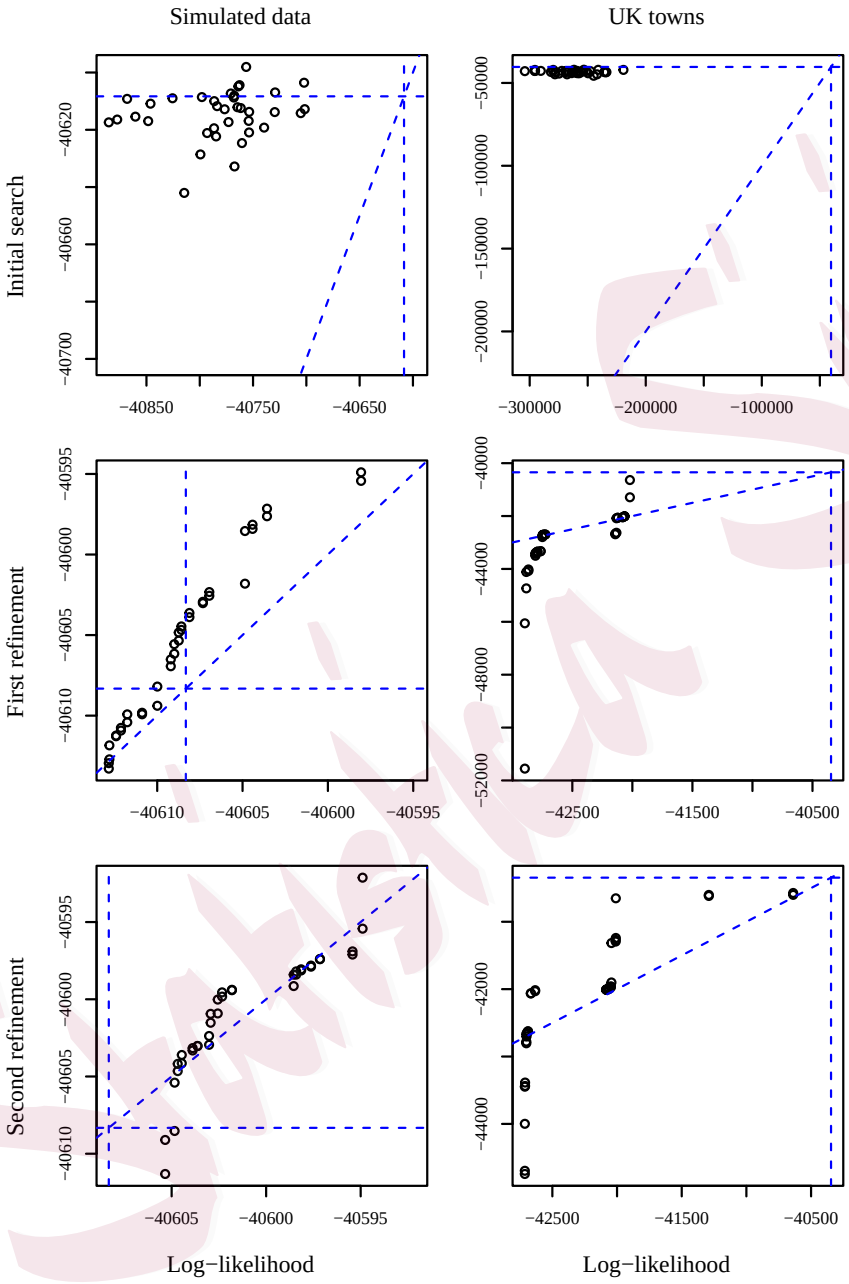

Figure 4: Three steps of a likelihood search on simulated data (left panel, model $\mathcal{A}$) and UK measles data (right panel, model $\mathcal{B}$). Dashed lines parallel to the axes show the log-likelihood at the truth (left panel) and the He et al. (2010) value (right panel). Points above the diagonal dashed line show improvement owing to the search step.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

simulated data (the first column) and $\mathcal{B}$ for the UK measles data (the second column). For now, we focus on the first column. The first row plots the log-likelihood obtained after the initial search against the log-likelihood of the randomly selected starting value. The horizontal and vertical dashed lines denote the likelihood at the truth, and the diagonal dashed line represents equality, so that points above the diagonal show improvement after the search. We find that IBPF robustly and rapidly approaches a neighborhood of the MLE, as measured by likelihood. The second row shows that further investigation of more successful searches can reliably obtain likelihood values higher than those at the truth. However, for a Monte Carlo search based on a Monte Carlo likelihood evaluation, it may be problematic to pinpoint the exact maximum in a high-dimensional space. The third row shows that continued searching does not lead to substantially better outcomes. When using these methods, we emphasize the need to make proper inferences despite imperfect maximization (Ionides et al., 2017; Ning et al., 2021).

The results in Figure 3 and the first column of Figure 4 show that IBPF can be effective for simulated data, but do not guarantee comparable performance for our data analysis. Indeed, model misspecification, which is inevitable for data analysis, may be expected to add difficulties to filtering

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

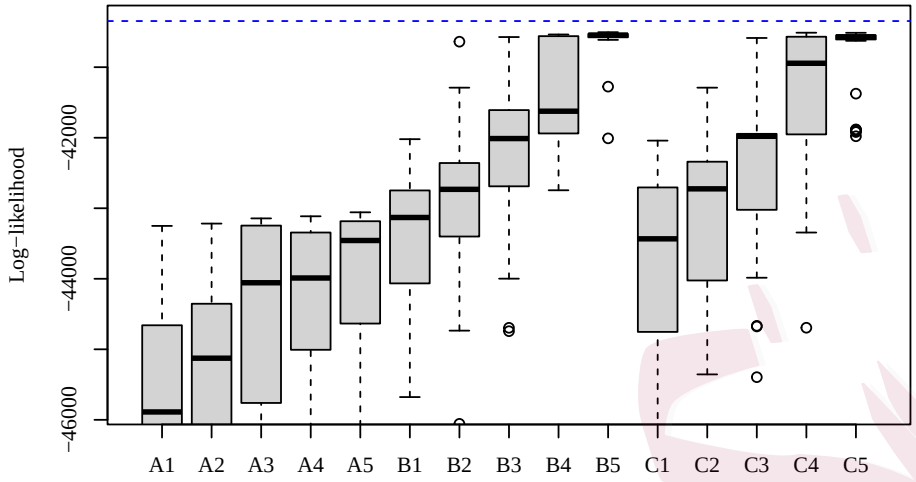

Figure 5: Fitting different models to the UK measles data using the method tested on simulated data. (A) 4×20 unit-specific parameters and nine shared parameters; (B) 13×20 unit-specific parameters and no shared parameters; (C) all unit-specific, but with immigration rather than coupling. The horizontal dashed line is the likelihood from He et al. (2010).

and therefore to numerical methods based on filtering. Rather, we view the simulation study as a lower bound on the effort required to perform effective inference on data, and therefore a starting point for investigating how to proceed with a data analysis. Before moving on to the data analysis, we briefly describe the details of the model.

3.3 IBPF applied to data

Figure 5 shows the results from fitting models $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ to the UK measles data using successive rounds of IBPF, applying the same method used for the simulation study shown in Figure 3. The dashed lines show the

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

sum of the log-likelihoods obtained by He et al. (2010), $\ell_{\mathrm{He}} = -40345.7$. The value ℓ_{He} corresponds to model $\mathcal{C}$. BPF with each town forming a separate block, applied to an uncoupled model such as $\mathcal{C}$, is equivalent to carrying out independent particle filters for each unit. Indeed, if we assign the published parameter values from He et al. (2010) to the `spatPomp` object for model $\mathcal{C}$, and apply the `bpfilter` function to carry out BPF, we retrieve ℓ_{He}, up to a Monte Carlo error. Models $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ in Figure 5 each show a shortfall relative to ℓ_{He}.

The largest shortfall is for $\mathcal{A}$, perhaps because $\mathcal{A}$ has more shared parameters than the evidence in the data supports. The results of He et al. (2010) suggest that the data are explained better when various parameters are a function of the town population. However, determining suitable functional forms for this relationship, and establishing regularities across towns that can be represented by shared parameters, remains an open problem. This may be investigated using panel methods, such as PanelPOMP models (Bretó et al., 2019), in addition to consideration of SpatPOMP models.

Models $\mathcal{B}$ and $\mathcal{C}$ yield comparable likelihoods, which is in contrast to the results shown in Figure 3. Fitting to simulated data from $\mathcal{B}$ (in the special case in which all unit-specific parameters are equal across all units), the shortfall for $\mathcal{C}$ in Figure 3 indicates that we obtain a substantially

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

worse fit when we approximate coupling using an uncoupled reservoir of infection. If the actual data were also explained substantially better by the coupled model, we would expect to see comparable results in the data analysis. However, because we do not, we conclude that this coupled model does not provide a substantially better explanation than the uncoupled model. Although various other candidate coupling mechanisms have been proposed (Bjørnstad et al., 2019), they have not yet been fitted to the full data, suggesting a lack of appropriate methodology to do so. Instead, Bjørnstad et al. (2019) consider summary statistics based on local fade-outs and re-introductions.

Figure 5 shows that model $\mathcal{B}$ has a small, but distinct shortfall compared with ℓ_{He}. For the simulated data, we do not observe this shortfall, and thus we deduce that the real data provide a more challenging optimization environment. When performing difficult optimization problems, it may be possible to develop helpful strategies specific to the model and data in question. For example, one could try merging unit-specific parameters from different searches, using the likelihood component for each unit to assess successful choices. However, optimization heuristics such as this do not have general theoretical support; if they obtain higher likelihoods, that is sufficient justification. For analyses of panel time series data using

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

PanelPOMP models, such methods have clearer theoretical support and have been found to be useful (Bretó et al., 2019).

4. Discussion

Theoretical interest in BPF algorithms has been inspired by the work of Rebeschini and Van Handel (2015), who showed mathematically that a BPF enjoys linear scaling properties under suitable conditions. However, they expressed pessimism about the practical applicability of their algorithm, which may help to explain its limited practical use. A similar algorithm was proposed independently under the name of a factored particle filter by Ng et al. (2002), who offered an empirical justification. However, the latter algorithm has also seen only limited use. The numerical results in Figure 3 of Ionides et al. (2021) suggest that BPF is particularly well suited to metapopulation models, in which we expect most of the population dynamics to occur at a local level, among individuals at one spatial unit. As a broad generalization, in ecological systems, the dispersion of individuals between spatial units is rare but dynamically important. Therefore, edge effects between blocks, which may be a serious problem for a BPF in a system with stronger spatial coupling (Ionides et al., 2021) is a relatively minor concern in metapopulation models. Blocks of size one unit are there-

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

fore a natural choice for block filtering of metapopulation models when the constituent populations are spatially distinct.

We have proposed a likelihood-maximization approach to inference. Much research has been done on inference for high-dimensional partially observed stochastic dynamic systems, and we have cited only the most directly relevant work. A Bayesian inference approach based on the ensemble Kalman filter is proposed by Katzfuss et al. (2020). An expectation-maximization approach based on a BPF is presented by Finke and Singh (2017). Spatiotemporal models with a convenient factorization across units are studied by Beskos et al. (2017) and Xu and Jasra (2019).

We have demonstrated a workflow that led to a likelihood-based assessment of measles metapopulation models ($\mathcal{A}$ and $\mathcal{B}$), with the possibility of finding evidence that they outperform the uncoupled model, $\mathcal{C}$. Our methodology successfully refutes $\mathcal{C}$ on simulated data, when the truth is within $\mathcal{A}$ and $\mathcal{B}$. However, there was no evidence of an advantage for these coupled models when the same comparison was carried out on the data. Likelihood maximization for the measles data fits the *common task framework* described by Donoho (2017, Section 6), with the likelihood value for $\mathcal{C}$ obtained by He et al. (2010) providing a benchmark challenge. Future improvements in models, perhaps facilitated by the open-source models

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

and methods accompanying this article, may obtain metapopulation models that convincingly beat model $\mathcal{C}$.

The computation time for Figures 3 and 5 was approximately 24hr on a single core of a computing node, for 4000 particles iterated 100 times over 730 time points (weekly data for 14 years) for 20 cities. Each box in these figures involves 36 replications, and thus took 24 hours on all cores of one node on a computing cluster. A measure of computational efficiency is the size of a problem that can be solved on this timescale, as discussed further in Section S3. Tasks that are considerably larger, perhaps 10^3 or more spatial units, may require additional approximations, such as those inherent in the ensemble Kalman filter (Evensen, 2009; Katzfuss et al., 2020) or other numerical filtering techniques (Whitehouse et al., 2022). However, we anticipate that many practical metapopulation analyses can be addressed within the scope we have demonstrated. The majority of the computational effort lies in simulating from the dynamic model. In the `spatPomp` R package, as in `pomp`, the user supplies a snippet of C code to simulate a single particle between a single arbitrary pair of times, and the package provides a vectorized form of this computation, which is used by the inference methodologies within the package. As a result, the computational performance is competitive even though the majority of the package is written in R (FitzJohn

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

et al., 2020, Table 1).

Supplementary material

Supplementary text presents the following: S1, parameters for the measles model; S2, algorithmic parameters and transformations; S3, computational efficiency; S4, varying the spatial autoregression parameter, r. Source code is at `https://github.com/ionides/ibpf_article`.

Acknowledgments

This work was supported by National Science Foundation grants DMS-1761603 and DMS 1646108, and a Texas A&M University Seed Fund Grant. We acknowledge an anonymous referee from a previous article who recommended implementing a block particle filter. We are also grateful for constructive feedback from two anonymous referees.

References

Asfaw, K., Ionides, E. L., and King, A. A. (2021). `spatPomp`: R package for statistical inference for spatiotemporal partially observed Markov processes. `https://cran.r-project.org/web/packages/spatPomp`.

Asfaw, K., Park, J., King, A. A., and Ionides, E. L. (2023). Statistical

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

inference for spatiotemporal partially observed Markov processes via the R package spatpomp. *arXiv:2101.01157v3*.

Bartlett, M. S. (1957). Measles periodicity and community size. *Journal of the Royal Statistical Society. Series A (General)*, 120(1):48–70.

Becker, A. D., Zhou, S. H., Wesolowski, A., and Grenfell, B. T. (2020). Coexisting attractors in the context of cross-scale population dynamics: Measles in London as a case study. *Proceedings of the Royal Society of London, Series B*, 287(1925):20191510.

Bengtsson, T., Bickel, P., and Li, B. (2008). Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems. In Speed, T. and Nolan, D., editors, *Probability and Statistics: Essays in Honor of David A. Freedman*, pages 316–334. Institute of Mathematical Statistics, Beachwood, OH.

Beskos, A., Crisan, D., Jasra, A., Kamatani, K., and Zhou, Y. (2017). A stable particle filter for a class of high-dimensional state-space models. *Advances in Applied Probability*, 49(1):24–48.

Bjørnstad, O. N. and Grenfell, B. T. (2001). Noisy clockwork: Time series analysis of population fluctuations in animals. *Science*, 293:638–643.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

Bjørnstad, O. N., Grenfell, B. T., Viboud, C., and King, A. A. (2019). Comparison of alternative models of human movement and the spread of disease. *BioRxiv:2019.12.19.882175*.

Bretó, C., He, D., Ionides, E. L., and King, A. A. (2009). Time series analysis via mechanistic models. *Annals of Applied Statistics*, 3:319–348.

Bretó, C. and Ionides, E. L. (2011). Compound Markov counting processes and their applications to modeling infinitesimally over-dispersed systems. *Stochastic Processes and their Applications*, 121:2571–2591.

Bretó, C., Ionides, E. L., and King, A. A. (2019). Panel data analysis via mechanistic models. *Journal of the American Statistical Association*, 115:1178–1188.

Cauchemez, S., Valleron, A., Boëlle, P., Flahault, A., and Ferguson, N. M. (2008). Estimating the impact of school closure on influenza transmission from sentinel data. *Nature*, 452:750–754.

Donoho, D. (2017). 50 years of data science. *Journal of Computational and Graphical Statistics*, 26(4):745–766.

Doucet, A., Jacob, P. E., and Rubenthaler, S. (2013). Derivative-free esti-

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

mation of the score vector and observed information matrix with application to state-space models. *arXiv:1304.5768*.

Doucet, A. and Johansen, A. (2011). A tutorial on particle filtering and smoothing: Fifteen years later. In Crisan, D. and Rozovsky, B., editors, *Oxford Handbook of Nonlinear Filtering*. Oxford University Press.

Earn, D. J. D., Rohani, P., Bolker, B. M., and Grenfell, B. T. (2000). A simple model for complex dynamical transitions in epidemics. *Science*, 287:667–670.

Eggo, R. M., Cauchemez, S., and Ferguson, N. M. (2011). Spatial dynamics of the 1918 influenza pandemic in England, Wales and the United States. *Journal of the Royal Society Interface*, 8(55):233–243.

Evensen, G. (2009). *Data assimilation: The ensemble Kalman filter*. Springer Science & Business Media.

Fine, P. E. and Clarkson, J. A. (1982). Measles in England and Wales—I: an analysis of factors underlying seasonal patterns. *International Journal of Epidemiology*, 11(1):5–14.

Finke, A. and Singh, S. S. (2017). Approximate smoothing and parameter

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

estimation in high-dimensional state-space models. *IEEE Transactions on Signal Processing*, 65(22):5982–5994.

FitzJohn, R. G., Knock, E. S., Whittles, L. K., Perez-Guzman, P. N., Bhatia, S., Guntoro, F., Watson, O. J., Whittaker, C., Ferguson, N. M., Cori, A., Baguelin, M., and Lees, J. A. (2020). Reproducible parallel inference and simulation of stochastic state space models using odin, dust, and mcstate. *Wellcome Open Research*, 5.

He, D., Ionides, E. L., and King, A. A. (2010). Plug-and-play inference for disease dynamics: Measles in large and small towns as a case study. *Journal of the Royal Society Interface*, 7:271–283.

Ionides, E. L., Asfaw, K., Park, J., and King, A. A. (2021). Bagged filters for partially observed interacting systems. *Journal of the American Statistical Association*, pre-published online.

Ionides, E. L., Bhadra, A., Atchadé, Y., and King, A. A. (2011). Iterated filtering. *Annals of Statistics*, 39:1776–1802.

Ionides, E. L., Bretó, C., and King, A. A. (2006). Inference for nonlinear dynamical systems. *Proceedings of the National Academy of Sciences of the USA*, 103:18438–18443.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

Ionides, E. L., Breto, C., Park, J., Smith, R. A., and King, A. A. (2017). Monte Carlo profile confidence intervals for dynamic systems. *Journal of the Royal Society Interface*, 14:1–10.

Ionides, E. L., Nguyen, D., Atchadé, Y., Stoev, S., and King, A. A. (2015). Inference for dynamic and latent variable models via iterated, perturbed Bayes maps. *Proceedings of the National Academy of Sciences of the USA*, 112(3):719—-724.

Kantas, N., Doucet, A., Singh, S. S., and Maciejowski, J. M. (2015). On particle methods for parameter estimation in state-space models. *Statistical Science*, 30(3):328–351.

Katzfuss, M., Stroud, J. R., and Wikle, C. K. (2020). Ensemble Kalman methods for high-dimensional hierarchical dynamic space-time models. *Journal of the American Statistical Association*, 115(530):866–885.

King, A. A., Nguyen, D., and Ionides, E. L. (2016). Statistical inference for partially observed Markov processes via the R package pomp. *Journal of Statistical Software*, 69:1–43.

Kitagawa, G. (1998). A self-organising state-space model. *Journal of the American Statistical Association*, 93:1203–1215.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

Lele, S. R., Dennis, B., and Lutscher, F. (2007). Data cloning: Easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. *Ecology Letters*, 10:551–563.

Lele, S. R., Nadeem, K., and Schmuland, B. (2010). Estimability and likelihood inference for generalized linear mixed models using data cloning. *Journal of the American Statistical Association*, 105:1617–1625.

Li, R., Pei, S., Chen, B., Song, Y., Zhang, T., Yang, W., and Shaman, J. (2020). Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science*, 368(6490):489–493.

Liu, J. and West, M. (2001). Combining parameter and state estimation in simulation-based filtering. In Doucet, A., de Freitas, N., and Gordon, N. J., editors, *Sequential Monte Carlo Methods in Practice*, pages 197–224. Springer, New York.

Manoli, G., Rossi, M., Pasetto, D., Deiana, R., Ferraris, S., Cassiani, G., and Putti, M. (2015). An iterative particle filter approach for coupled hydro-geophysical inversion of a controlled infiltration experiment. *Journal of Computational Physics*, 283:37–51.

Ng, B., Peshkin, L., and Pfeffer, A. (2002). Factored particles for scal-

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

able monitoring. *Proceedings of the 18th Conference on Uncertainty and Artificial Intelligence*, pages 370–377.

Nguyen, D. and Ionides, E. L. (2015). A second-order iterated smoothing algorithm. *Statistics & Computing*, 27(6):1677–1692.

Ning, N. and Ionides, E. L. (2023). Iterated block particle filter for high-dimensional parameter learning: Beating the curse of dimensionality. *Journal of Machine Learning Research*, to appear.

Ning, N., Ionides, E. L., and Ritov, Y. (2021). Scalable Monte Carlo inference and rescaled local asymptotic normality. *Bernoulli*, 27:2532–2555.

Park, J. and Ionides, E. L. (2020). Inference on high-dimensional implicit dynamic models using a guided intermediate resampling filter. *Statistics & Computing*, 30:1497–1522.

Rebeschini, P. and Van Handel, R. (2015). Can local particle filters beat the curse of dimensionality? *The Annals of Applied Probability*, 25(5):2809–2866.

Snyder, C., Bengtsson, T., and Morzfeld, M. (2015). Performance bounds for particle filters using the optimal proposal. *Monthly Weather Review*, 143(11):4750–4761.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0188

Whitehouse, M., Whiteley, N., and Rimella, L. (2022). Consistent and fast inference in compartmental models of epidemics using Poisson Approximate Likelihoods. *arXiv:2205.13602*.

Xia, Y., Bjørnstad, O. N., and Grenfell, B. T. (2004). Measles metapopulation dynamics: A gravity model for epidemiological coupling and dynamics. *American Naturalist*, 164(2):267–281.

Xu, Y. and Jasra, A. (2019). Particle filters for inference of high-dimensional multivariate stochastic volatility models with cross-leverage effects. *Foundations of Data Science*, 1(1):61.

Statistica Sinica