

Bagged filters for partially observed spatiotemporal systems

Edward L. Ionides*

Department of Statistics, University of Michigan

Kidus Asfaw

Department of Statistics, University of Michigan

Joonha Park

Department of Mathematics & Statistics, Boston University

Aaron A. King

Department of Ecology & Evolutionary Biology, University of Michigan

Abstract

Bagging (i.e., bootstrap aggregating) involves combining an ensemble of bootstrap estimators. The particle filter, also known as sequential Monte Carlo or the bootstrap filter, is a widely used algorithm for estimating the latent states or likelihood of a partially observed Markov process. Our bagged filter (BF) methodology combines an ensemble of particle filters, using spatiotemporally localized weights to select successful filters at each point of space and time. In some situations, BF methodology can theoretically beat a curse of dimensionality affecting standard particle filters; in others, BF fails in theory to beat the curse, but nevertheless evinces practical advantages. Our focus is on evaluation of the likelihood function, though BF theory and methodology are also pertinent to latent state estimation. Applications suited to BF methodology include spatiotemporal analysis of epidemiological and ecological systems. We show that BF methodology can out-perform an ensemble Kalman filter on a coupled population dynamics model arising in the epidemiology of infectious disease. We also find that a block particle filter performs well on this task, though the bagged filter respects smoothness and conservation laws that a block particle filter can violate.

Keywords: Particle filter; Sequential Monte Carlo; Markov process; Population dynamics

*This work was supported by National Science Foundation grants DMS-1761603 and DMS-1646108, and National Institutes of Health grants 1-U54-GM111274 and 1-U01-GM110712.

1 Introduction

Particle filter (PF) algorithms are a widespread tool for inference on nonlinear partially observed dynamic systems (Doucet et al., 2001; Doucet and Johansen, 2011; Bretó et al., 2009; Ionides et al., 2015). Modern particle filtering traces back to Gordon et al. (1993) where the algorithm was called the *bootstrap filter*. Standard PF methods suffer from a curse of dimensionality (COD), defined as an exponential increase in computational demands as the problem size grows, which limits its applicability to spatiotemporal systems (Bengtsson et al., 2008; Snyder et al., 2015; Rebeschini and van Handel, 2015). The COD presents empirically as numerical instability of the Monte Carlo algorithm for attainable numbers of particles. Bootstrap aggregating (i.e., bagging) is a methodology to improve numerically unstable estimators by combining an ensemble of replicated bootstrap calculations (Breiman, 1996). We consider algorithms combining an ensemble of replicated particle filters, and we use the name *bagged filter*. The task of each bootstrap replicate is to provide some predictive skill at some points in space and time. The bagged filter then proceeds to combine these replicates, assessing which replicates are reliable predictors at a given spatiotemporal location.

Much previous research has investigated scalable approaches to filtering and inference with applications to spatiotemporal systems. Our bagged filters are in the class of *plug-and-play* algorithms, meaning that they require as input a simulator for the latent dynamic process but not an evaluator of transition probabilities (Bretó et al., 2009; He et al., 2010). This simulation-based approach, also known as *likelihood-free* (Brehmer et al., 2020) or *equation-free* (Kevrekidis and Samaey, 2009), facilitates application to a wide class of models. The ensemble Kalman filter (Evensen and van Leeuwen, 1996; Katzfuss et al., 2020) is a widely used plug-and-play method which uses simulations to parameterize a Gaussian-inspired filtering rule. Another plug-and-play approach to combat the COD is the block particle filter (Rebeschini and van Handel, 2015; Ng et al., 2002). Both ensemble Kalman filter and block particle filter methods construct trajectories that can violate smoothness and conservation properties of the dynamic model. By contrast, our bagged filters are built using valid trajectories of the dynamic model, but make localization approximations when comparing these

trajectories to data. It is an empirical question how these strengths and weaknesses play out for a particular data analysis. The plug-and-play property of an algorithm makes a general numerical implementation possible: we provide a software environment within which alternative methods can be applied.

The replicated stochastic trajectories in a bagged filter form an ensemble of representations of the dynamic system. Unlike ensemble Kalman filters (Lei et al., 2010) the bagged replicates are independent in a Monte Carlo sense. Bagged filters therefore bear some resemblance to *poor man’s ensemble* forecasting methodology in which a collection of independently constructed forecasts is generated using different models and methods (Ebert, 2001). Poor man’s ensembles have sometimes been found to have greater forecasting skill than any one forecast (Leutbecher and Palmer, 2008; Palmer, 2002; Chandler, 2013). One explanation for this phenomenon is that even a hypothetically perfect model cannot filter and forecast well using methodology afflicted by the COD. We show that bagged filter methodology can relieve this limitation. From this perspective, the independence of the forecasts in the poor man’s ensemble, rather than the diversity of model structures, may be the key to its success.

In a basic bagged filter, each replicate simply simulates a realization of the latent process model, corresponding to a degenerate instance of PF with only a single particle. We call this the unadapted bagged filter (UBF) since each replicate in the ensemble does not track the data. While formally beating the COD under a weak mixing assumption, UBF can have poor numerical behavior if a very large number of replicates are needed to reach this happy asymptotic limit. Our empirical results show that UBF may indeed be a useful algorithm in some situations. More importantly, however, it provides a route to two other algorithms, which we call the adapted bagged filter (ABF) and adapted bagged filter with intermediate resampling (ABF-IR). ABF and ABF-IR carry out some resampling on each replicate and therefore enable each replicate to track the data, in a weak sense, while resisting the COD. This comes at a price: ABF no longer asymptotically beats COD, and ABF-IR does so only under restrictive conditions. However, we demonstrate useful behavior for these algorithms on finite samples.

To set up notation, we suppose the latent process is comprised of a collection of units, labeled $\{1, 2, \dots, U\}$ which we write as $1:U$. An unobserved, latent Markov process at a discrete sequence of observation times is written as $\{\mathbf{X}_n, n \in 0:N\}$, with $\mathbf{X}_n = X_{1:U,n}$ taking values in \mathbb{X}^U . The initial value \mathbf{X}_0 may be stochastic or deterministic. Observations are made on each unit, modeled by an observable process $\{\mathbf{Y}_n = Y_{1:U,n}, n \in 1:N\}$ which takes values in \mathbb{Y}^U . Observations are modeled as being conditionally independent given the latent process. The conditional independence of measurements applies over both space and time, so $Y_{u,n}$ is conditionally independent of $\{X_{\tilde{u},\tilde{n}}, Y_{\tilde{u},\tilde{n}}, (\tilde{u},\tilde{n}) \neq (u,n)\}$ given $X_{u,n}$. This unit structure for the observation process is not necessary for all that follows (see Sec. S1). We suppose the existence of a joint density $f_{\mathbf{X}_{0:N}, \mathbf{Y}_{1:N}}$ of $\{X_{1:U,n}\}$ and $\{Y_{1:U,n}\}$ with respect to some appropriate measure. We use the same subscript notation to denote other marginal and conditional densities. The data are $y_{u,n}$ for spatial unit u at time n . For a fixed value of U , this model fits into the partially observed Markov process (POMP) framework (Bretó et al., 2009), also known as a state space model or hidden Markov model. Spatiotemporal POMP models are characterized by a latent Markov process with dimension proportional to U . We use the term SpatPOMP to describe POMP models indexed by u as well as n . In accordance with the spatial interpretation of the units $u \in 1:U$, we may suppose that units sufficiently distant in some sense approach independence, a phenomenon called *weak coupling*. It is worth pointing out that inference techniques designed for SpatPOMPs are also relevant in high-dimensional stochastic dynamic systems where u does not index spatial locations. In the following, we use the ordering on the set of observations corresponding to their labels $1:U$ to define the set of observations preceding unit u at time n as

$$A_{u,n} = \{(\tilde{u}, \tilde{n}) : 1 \leq \tilde{n} < n \text{ or } (\tilde{n} = n \text{ and } \tilde{u} < u)\}. \quad (1)$$

The ordering of the spatial locations in (1) might seem artificial, and indeed densities such as $f_{X_{u,n}|X_{A_{u,n}}}$ will frequently be hard to compute or simulate from. However, the algorithms we study do not carry out such computations, but only evaluate the measurement model on these neighborhoods. To represent weak coupling, we suppose there is a neighborhood $B_{u,n} \subset A_{u,n}$ such that the latent process on $A_{u,n} \setminus B_{u,n}$ is approximately conditionally independent of $X_{u,n}$

given data on $B_{u,n}$.

The inputs for the UBF, ABF and ABF-IR are listed in Table 1. The plug-and-play property is evident since all three algorithms require as input a simulator for the latent dynamic process but not an evaluator of transition probabilities. Our primary interest is in the estimation of the log likelihood for the data given the model, $\ell = \log f_{\mathbf{Y}_{1:N}}(\mathbf{y}_{1:N}^*)$, which is of fundamental importance in both Bayesian and non-Bayesian statistical inference. Monte Carlo estimates of the log likelihood underpin modern methods for full-information statistical inference on partially observed dynamic systems (Andrieu et al., 2010; Ionides et al., 2015, 2017). Filtering theory ties together the three tasks of likelihood evaluation, forecasting, and inferring latent dynamic variables (Crisan and Rozovskii, 2011). Successful likelihood evaluation is therefore closely related to success at forecasting and latent process reconstruction.

The pseudocodes for the algorithms that follow will adopt a convention that implicit loops are carried out over all free indices. For example, the construction of $w_{u,n,i}^P$ in UBF has an implicit loop over u , n and i . The construction of $w_{u,n,i,j}^P$ in ABF has an implicit loop over u , i and j but not over n since n is already defined in an explicit loop and is therefore not a free index. Table 1 lists the ranges of the implicit loops for all following algorithms.

Table 1. Bagged filter inputs, outputs and implicit loops.

input:

simulator for $f_{\mathbf{X}_0}(\mathbf{x}_0)$ and $f_{\mathbf{X}_n|\mathbf{X}_{n-1}}(\mathbf{x}_n | \mathbf{x}_{n-1})$

evaluator for $f_{Y_{u,n}|X_{u,n}}(y_{u,n} | x_{u,n})$

number of replicates, \mathcal{I}

neighborhood structure, $B_{u,n}$

data, $\mathbf{y}_{1:N}^*$

ABF and ABF-IR: particles per replicate, J

ABF-IR: number of intermediate timesteps, S

ABF-IR: measurement variance parameterizations, $\overleftarrow{\mathbf{v}}_{u,n}$ and $\overrightarrow{\mathbf{v}}_{u,n}$

ABF-IR: approximate process and observation mean functions, $\boldsymbol{\mu}$ and $h_{u,n}$
output:

Log likelihood estimate, $\ell^{\text{MC}} = \sum_{n=1}^N \sum_{u=1}^U \ell_{u,n}^{\text{MC}}$
implicit loops:
 u in $1:U$, n in $1:N$, i in $1:\mathcal{I}$, j in $1:J$

The UBF algorithm that follows amounts to a bagged filter with a single particle per replicate, with no weighting or resampling within each replicate.

UBF. Unadapted bagged filter.

Simulate $\mathbf{X}_{0:N,i} \sim f_{\mathbf{X}_{0:N}}(\mathbf{x}_{0:N})$

Measurement weights, $w_{u,n,i}^M = f_{Y_{u,n}|X_{u,n}}(y_{u,n}^* | X_{u,n,i})$

Prediction weights, $w_{u,n,i}^P = \prod_{(\tilde{u}, \tilde{n}) \in B_{u,n}} w_{\tilde{u}, \tilde{n}, i}^M$
 $\ell_{u,n}^{\text{MC}} = \log \left(\sum_{i=1}^{\mathcal{I}} w_{u,n,i}^M w_{u,n,i}^P \right) - \log \left(\sum_{i=1}^{\mathcal{I}} w_{u,n,i}^P \right)$

Theorem 1 shows that UBF can beat COD. However, an algorithm such as UBF, which includes no resampling, will fail on long time series absent rapid spatiotemporal mixing, which allows simulated particles to remain relevant over the course of a long time series. It may in practice be necessary to select simulations consistent with the data, much as standard PF algorithms do. We look for approaches that build on the basic insight of UBF while having superior practical performance.

Whereas the full global filtering problem may be intractable via importance sampling methods, a version of this problem local in space and time may nevertheless be feasible. Drawing a sample from the conditional density, $f_{\mathbf{X}_n|\mathbf{Y}_n, \mathbf{X}_{n-1}}$, is distinct from the filtering problem: we call this operation *adapted simulation*. For models where \mathbf{X}_n is close to \mathbf{X}_{n-1} , adapted simulation requires a local calculation that may be much easier than the full filtering calculation. The following adapted bagged filter (ABF) is constructed under a hypothesis that the adapted simulation problem is tractable. The adapted simulations are then reweighted in a neighborhood of each point in space and time to construct a local approximation to the filtering problem which leads to an estimate of the likelihood.

ABF. Adapted bagged filter.

Initialize adapted simulation: $\mathbf{X}_{0,i}^A \sim f_{\mathbf{X}_0}(\mathbf{x}_0)$

For n in $1:N$

Proposals: $\mathbf{X}_{n,i,j}^P \sim f_{\mathbf{X}_n|X_{1:U,n-1}}(\mathbf{x}_n | \mathbf{X}_{n-1,i}^A)$

Measurement weights: $w_{u,n,i,j}^M = f_{Y_{u,n}|X_{u,n}}(y_{u,n}^* | X_{u,n,i,j}^P)$

Adapted resampling weights: $w_{n,i,j}^A = \prod_{u=1}^U w_{u,n,i,j}^M$

Resampling: $\mathbb{P}[r(i) = a] = w_{n,i,a}^A \left(\sum_{k=1}^J w_{n,i,k}^A \right)^{-1}$

$\mathbf{X}_{n,i}^A = \mathbf{X}_{n,i,r(i)}^P$

$w_{u,n,i,j}^P = \prod_{\tilde{n}=1}^{n-1} \left[\frac{1}{J} \sum_{k=1}^J \prod_{\tilde{u}: (\tilde{u}, \tilde{n}) \in B_{u,n}} w_{\tilde{u}, \tilde{n}, i, k}^M \right] \prod_{\tilde{u}: (\tilde{u}, n) \in B_{u,n}} w_{\tilde{u}, n, i, j}^M$

End for

$$\ell_{u,n}^{\text{MC}} = \log \left(\frac{\sum_{i=1}^{\mathcal{I}} \sum_{j=1}^J w_{u,n,i,j}^M w_{u,n,i,j}^P}{\sum_{i=1}^{\mathcal{I}} \sum_{j=1}^J w_{u,n,i,j}^P} \right)$$

In addition, we consider an algorithm which adds intermediate resampling to the adapted simulation replicates, and is therefore termed ABF-IR. Intermediate resampling involves assessing the satisfactory progress of particles toward the subsequent observation at a collection of times between observations. This is well defined when the latent system is a continuous time process $\{\mathbf{X}(t)\}$ with observation times $t_{1:N}$ such that $\mathbf{X}_n = \mathbf{X}(t_n)$. For a continuous time latent process model, intermediate resampling can have favorable scaling properties

when the intermediate times

$$t_{n-1} = t_{n,0} < t_{n,1} < \dots < t_{n,S} = t_n$$

scale with $S \approx U$ (Park and Ionides, 2020). In the case $S = 1$, ABF-IR reduces to ABF. Intermediate resampling was developed in the context of sequential Monte Carlo (Del Moral and Murray, 2015; Park and Ionides, 2020); however, the same theory and methodology can be applied to the simpler and easier problem of adapted simulation. ABF-IR employs a guide function to gauge the compatibility of each particle with future data. This is a generalization of the popular auxiliary particle filter (Pitt and Shepard, 1999). The implementation in this pseudocode constructs the guide $g_{n,s,i,j}$ using a simulated moment method proposed by Park and Ionides (2020). The quantities $\mathbf{X}_{n,i,j}^G$, $V_{u,n,i}$, $\boldsymbol{\mu}_{n,s,i,j}^{\text{IP}}$, $V_{u,n,s,i,j}^{\text{meas}}$, $V_{u,n,s,i}^{\text{proc}}$ and $\theta_{u,n,s,i,j}$ constructed in ABF-IR are used only to construct $g_{n,s,i,j}$. Heuristically, we use guide simulations to approximate the variance of the increment in each particle between time points, and we augment the measurement variance to account for both dynamic variability and measurement error. The guide function affects numerical performance of the algorithm but not its correctness: it enables a computationally convenient approximation to improve performance on the intractable target problem. Our guide function supposes the availability of a deterministic function approximating evolution of the mean of the latent process, written as

$$\boldsymbol{\mu}(\mathbf{x}, s, t) \approx \mathbb{E}[\mathbf{X}(t) \mid \mathbf{X}(s) = \mathbf{x}].$$

Further, the guide requires that the measurement model has known conditional mean and variance as a function of the model parameter vector θ , written as

$$\begin{aligned} h_{u,n}(x_{u,n}) &= \mathbb{E}[Y_{u,n} \mid X_{u,n} = x_{u,n}] \\ \vec{v}_{u,n}(x_{u,n}, \theta) &= \text{Var}(Y_{u,n} \mid X_{u,n} = x_{u,n}; \theta) \end{aligned}$$

Also required for ABF-IR is an inverse function $\overleftarrow{v}_{u,n}$ such that

$$\vec{v}_{u,n}(x_{u,n}, \overleftarrow{v}_{u,n}(V, x_{u,n}, \theta)) = V.$$

ABF-IR. Adapted bagged filter with intermediate resampling.

Initialize adapted simulation: $\mathbf{X}_{0,i}^A \sim f_{\mathbf{X}_0}(\mathbf{x}_0)$

For n in $1:N$

Guide simulations: $\mathbf{X}_{n,i,j}^G \sim f_{\mathbf{X}_n|\mathbf{X}_{n-1}}(\mathbf{x}_n | \mathbf{X}_{n-1,i}^A)$

Guide sample variance: $V_{u,n,i} = \text{Var}\{h_{u,n}(X_{u,n,i,j}^G), j \text{ in } 1:J\}$

$g_{n,0,i,j}^R = 1$ and $\mathbf{X}_{n,0,i,j}^{\text{IR}} = \mathbf{X}_{n-1,i}^A$

For s in $1:S$

Intermediate proposals: $\mathbf{X}_{n,s,i,j}^{\text{IP}} \sim f_{\mathbf{X}_{n,s}|\mathbf{X}_{n,s-1}}(\cdot | \mathbf{X}_{n,s-1,i,j}^{\text{IR}})$

$\boldsymbol{\mu}_{n,s,i,j}^{\text{IP}} = \boldsymbol{\mu}(\mathbf{X}_{n,s,i,j}^{\text{IP}}, t_{n,s}, t_n)$

$V_{u,n,s,i,j}^{\text{meas}} = \vec{v}_u(\theta, \mu_{u,n,s,i,j}^{\text{IP}})$

$V_{u,n,s,i}^{\text{proc}} = V_{u,n,i}(t_n - t_{n,s}) / (t_n - t_{n,0})$

$\theta_{u,n,s,i,j} = \overleftarrow{v}_u(V_{u,n,s,i,j}^{\text{meas}} + V_{u,n,s,i}^{\text{proc}}, \mu_{u,n,s,i,j}^{\text{IP}})$

$g_{n,s,i,j} = \prod_{u=1}^U f_{Y_{u,n}|X_{u,n}}(y_{u,n}^* | \mu_{u,n,s,i,j}^{\text{IP}}; \theta_{u,n,s,i,j})$

Guide weights: $w_{n,s,i,j}^G = g_{n,s,i,j} / g_{n,s-1,i,j}^R$

Resampling: $\mathbb{P}[r(i,j) = a] = w_{n,s,i,a}^G \left(\sum_{k=1}^J w_{n,s,i,k}^G \right)^{-1}$

$\mathbf{X}_{n,s,i,j}^{\text{IR}} = \mathbf{X}_{n,s,i,r(i,j)}^{\text{IP}}$ and $g_{n,s,i,j}^R = g_{n,s,i,r(i,j)}$

End For

Set $\mathbf{X}_{n,i}^A = \mathbf{X}_{n,S,i,1}^{\text{IR}}$

Measurement weights: $w_{u,n,i,j}^M = f_{Y_{u,n}|X_{u,n}}(y_{u,n}^* | X_{u,n,i,j}^G)$

$w_{u,n,i,j}^P = \prod_{\tilde{n}=1}^{n-1} \left[\frac{1}{J} \sum_{a=1}^J \prod_{\tilde{u}: (\tilde{u}, \tilde{n}) \in B_{u,n}} w_{\tilde{u}, \tilde{n}, i, a}^M \right] \prod_{\tilde{u}: (\tilde{u}, n) \in B_{u,n}} w_{\tilde{u}, n, i, j}^M$

End for

$\ell_{u,n}^{\text{MC}} = \log \left(\frac{\sum_{i=1}^{\mathcal{I}} \sum_{j=1}^J w_{u,n,i,j}^M w_{u,n,i,j}^P}{\sum_{i=1}^{\mathcal{I}} \sum_{j=1}^J w_{u,n,i,j}^P} \right)$

This guide function is applicable to spatiotemporal versions of a broad range of population and compartment models used to model dynamic systems in ecology, epidemiology, and elsewhere. Other guide functions could be developed and inserted into the ABF-IR algorithm, including other constructions considered by Park and Ionides (2020). The number of particles for the guide simulations in ABF-IR does not necessarily have to be set equal to the number of intermediate proposals. We make that simplification here since there is no compelling reason against it.

One might wonder why it is appropriate to keep many particle representations at intermediate timesteps while resampling down to a single representative at each observation time. Part of the answer is that adaptive simulation can fail when one resamples down to a single particle too often (Sec. S2).

2 Likelihood factorizations and their approximations

The approximation error for the algorithms we study here can be divided into two sources: a localization bias due to conditioning on a finite neighborhood, and Monte Carlo error. The localization bias does not disappear in the limit as Monte Carlo effort increases. It does become small as the conditioning neighborhood increases, but the Monte Carlo effort grows exponentially in the size of this neighborhood. The class of problems for which these algorithms are useful are ones where a relatively small neighborhood is adequate. Although the filtering inference is carried out using localization, the simulation of the process is carried out globally which avoids the introduction of additional boundary effects and ensures that the simulations comply with any constraint properties of the global process simulator.

In this subsection, we consider the deterministic limit of each algorithm for infinite Monte Carlo effort. We explain why the UBF, ABF and ABF-IR algorithms approximately target the likelihood function, subject to suitable mixing behavior. Subsequently, we consider the scaling properties as Monte Carlo effort increases. We adopt a convention that densities involving $Y_{u,n}$ are implicitly evaluated at the data, $y_{u,n}^*$, and densities involving $X_{u,n}$ are implicitly evaluated at $x_{u,n}$ unless otherwise specified. We write $A_{u,n}^+ = A_{u,n} \cup (u, n)$ and

$B_{u,n}^+ = B_{u,n} \cup (u, n)$. The essential ingredient in all the algorithms is a localization of the likelihood, which may be factorized sequentially as

$$f_{Y_{1:U}, 1:N} = \prod_{n=1}^N \prod_{u=1}^U f_{Y_{u,n}|Y_{A_{u,n}}} = \prod_{n=1}^N \prod_{u=1}^U \frac{f_{Y_{A_{u,n}^+}}}{f_{Y_{A_{u,n}}}}.$$

In particular, the approximations assume that the full history $A_{u,n}$ can be well approximated by a neighborhood $B_{u,n} \subset A_{u,n}$. UBF approximates $f_{Y_{u,n}|Y_{A_{u,n}}}$ by

$$f_{Y_{u,n}|Y_{B_{u,n}}} = \frac{f_{Y_{B_{u,n}^+}}}{f_{Y_{B_{u,n}}}} = \frac{\int f_{Y_{B_{u,n}^+}|X_{B_{u,n}^+}} f_{X_{B_{u,n}^+}} dx_{B_{u,n}^+}}{\int f_{Y_{B_{u,n}}|X_{B_{u,n}}} f_{X_{B_{u,n}}} dx_{B_{u,n}}}.$$

For $B \subset 1:U \times 1:N$, define $B^{[m]} = B \cap (1:U \times \{m\})$. ABF and ABF-IR build on the following identity,

$$f_{Y_{A_{u,n}}} = \int f_{\mathbf{X}_0} \left[\prod_{m=1}^n f_{\mathbf{X}_m|\mathbf{X}_{m-1}, \mathbf{Y}_m} f_{Y_{A_{u,n}^{[m]}|\mathbf{X}_{m-1}}} \right] d\mathbf{x}_{0:n},$$

where $f_{\mathbf{X}_m|\mathbf{X}_{m-1}, \mathbf{Y}_m}$ is called the adapted transition density. The adapted process (i.e., a stochastic process following the adapted transition density) can be interpreted as a one-step greedy procedure using the data to guide the latent process. Let $g_{\mathbf{X}_{0:N}, \mathbf{X}_{1:N}^P}(\mathbf{x}_{0:N}, \mathbf{x}_{1:N}^P)$ be the joint density of the adapted process and the proposal process,

$$\begin{aligned} g_{\mathbf{X}_{0:N}, \mathbf{X}_{1:N}^P}(\mathbf{x}_{0:N}, \mathbf{x}_{1:N}^P) &= f_{\mathbf{X}_0}(\mathbf{x}_0) \times \\ &\prod_{n=1}^N f_{\mathbf{X}_n|\mathbf{X}_{n-1}, \mathbf{Y}_n}(\mathbf{x}_n | \mathbf{x}_{n-1}, \mathbf{y}_n^*) f_{\mathbf{X}_n|\mathbf{X}_{n-1}}(\mathbf{x}_n^P | \mathbf{x}_{n-1}). \end{aligned} \quad (2)$$

For $B \subset 1:U \times 1:N$, define $B^{[m]} = B \cap (1:U \times \{m\})$ and set

$$\gamma_B = \prod_{m=1}^N f_{Y_{B^{[m]}|\mathbf{X}_{m-1}}}(\mathbf{y}_{B^{[m]}}^* | \mathbf{X}_{m-1}),$$

using the convention that an empty density f_{Y_\emptyset} evaluates to 1. Denoting \mathbb{E}_g for expectation for $(\mathbf{X}_{0:N}, \mathbf{X}_{1:N}^P)$ having density $g_{\mathbf{X}_{0:N}, \mathbf{X}_{1:N}^P}$, we have

$$f_{Y_{u,n}|Y_{A_{u,n}}} = \frac{\mathbb{E}_g[\gamma_{A_{u,n}^+}]}{\mathbb{E}_g[\gamma_{A_{u,n}}]}.$$

Estimating this ratio by Monte Carlo sampling from g is problematic due to the growing size of $A_{u,n}$. Thus, ABF and ABF-IR make a localized approximation,

$$\frac{\mathbb{E}_g[\gamma_{A_{u,n}^+}]}{\mathbb{E}_g[\gamma_{A_{u,n}}]} \approx \frac{\mathbb{E}_g[\gamma_{B_{u,n}^+}]}{\mathbb{E}_g[\gamma_{B_{u,n}}]}. \quad (3)$$

The conditional log likelihood estimate $\ell_{u,n}^{\text{MC}}$ in ABF and ABF-IR come from replacing the expectations on the right hand side of (3) with averages over Monte Carlo replicates of simulations from the adapted process. To see that we expect the approximation in (3) to hold when dependence decays across spatiotemporal distance, we can write

$$\begin{aligned} \gamma_{A_{u,n}} &= \gamma_{B_{u,n}} \gamma_{B_{u,n}^c} \\ \gamma_{A_{u,n}^+} &= \gamma_{B_{u,n}^+} \gamma_{B_{u,n}^c}, \end{aligned}$$

where $B_{u,n}^c$ is the complement of $B_{u,n}$ in $A_{u,n}$. As long as $\gamma_{B_{u,n}^c}$ is approximately independent of $X_{u,n}$ under g , this term approximately cancels in the numerator and denominator of the right hand side of (3).

3 UBF theory

For each pair (U, N) we suppose there is a dataset $\mathbf{y}_{1:N}^*$ and a model $f_{\mathbf{X}_{0:N}, \mathbf{Y}_{1:N}}$. We are interested in results that hold for all (U, N) , but we do not require that the models for $(U_1, N_1) \neq (U_2, N_2)$ are nested in any way. The following conditions impose requirements that distant regions of space-time behave similarly and have only weak dependence. The conditions are written non-asymptotically in terms of $\epsilon > 0$ and $Q > 1$ which are used to bound the asymptotic bias and variance in Theorem 1. Stronger bounds are obtained when the conditions hold for small ϵ_{A1} , ϵ_{A4} and Q .

Assumption A1. *There is an $\epsilon_{A1} > 0$, independent of U and N , and a collection of neighborhoods $\{B_{u,n} \subset A_{u,n}, u \in 1:U, n \in 1:N\}$ such that, for all u and n , any bounded real-valued*

function $|h(x)| \leq 1$, and any value of $x_{B_{u,n}^c}$,

$$\left| \int h(x_{u,n}) f_{X_{u,n}|Y_{B_{u,n}}, X_{B_{u,n}^c}}(x_{u,n} | y_{B_{u,n}}^*, x_{B_{u,n}^c}) dx_{u,n} - \int h(x_{u,n}) f_{X_{u,n}|Y_{B_{u,n}}}(x_{u,n} | y_{B_{u,n}}^*) dx_{u,n} \right| < \epsilon_{A1}.$$

Assumption A2. For the collection of neighborhoods in Assumption A1, with $B_{u,n}^+ = B_{u,n} \cup (u, n)$, there is a constant b , depending on ϵ_{A1} but not on U and N , such that

$$\sup_{u \in 1:U, n \in 1:N} |B_{u,n}^+| \leq b.$$

Assumption A3. There is a constant Q , independent of U and N , such that, for all u and n ,

$$Q^{-1} < f_{Y_{u,n}|X_{u,n}}(y_{u,n}^* | x_{u,n}) < Q$$

Assumption A4. There exists $\epsilon_{A4} > 0$, independent of U and N , such that the following holds. For each u, n , a set $C_{u,n} \subset (1:U) \times (0:N)$ exists such that $(\tilde{u}, \tilde{n}) \notin C_{u,n}$ implies $B_{u,n}^+ \cap B_{\tilde{u}, \tilde{n}}^+ = \emptyset$ and

$$|f_{X_{B_{\tilde{u}, \tilde{n}}^+}|X_{B_{u,n}^+}} - f_{X_{B_{u,n}^+}}| < \epsilon_{A4} f_{X_{B_{\tilde{u}, \tilde{n}}^+}}$$

Further, there is a uniform bound $|C_{u,n}| \leq c$.

The two mixing conditions in Assumptions A1 and A4 are subtly different. Assumption A1 is a conditional mixing property, dependent on the data, whereas A4 asserts a form of unconditional mixing. Although both capture a similar concept of weak coupling, conditional and unconditional mixing properties do not readily imply one another. Assumption A3 is a compactness condition of a type that has proved useful in the theory of particle filters despite the rarity of its holding exactly. Theorem 1 shows that these conditions let UBF compute the likelihood with a Monte Carlo variance of order $UN\mathcal{I}^{-1}$ with a bias of order $UN\epsilon$.

Theorem 1. Let ℓ^{MC} denote the Monte Carlo likelihood approximation constructed by UBF. Consider a limit with a growing number of bootstrap replicates, $\mathcal{I} \rightarrow \infty$, and suppose assumptions A1, A2 and A3. There are quantities $\epsilon(U, N)$ and $V(U, N)$, with bounds $|\epsilon| < \epsilon_{A1} Q^2$ and $V < Q^{4b} U^2 N^2$, such that

$$\mathcal{I}^{1/2} [\ell^{MC} - \ell - \epsilon UN] \xrightarrow[\mathcal{I} \rightarrow \infty]{d} \mathcal{N}[0, V], \quad (4)$$

where $\xrightarrow[\mathcal{I} \rightarrow \infty]{d}$ denotes convergence in distribution and $\mathcal{N}[\mu, \Sigma]$ is the normal distribution with mean μ and variance Σ . If additionally Assumption A4 holds, we obtain an improved variance bound

$$V < Q^{4b} UN (c + \epsilon_{A4} (UN - c)). \quad (5)$$

Proof. A complete proof is given in Sec. S3. Briefly, the assumptions imply a multivariate central limit theorem for $\{\ell_{u,n}^{MC}, (u, n) \in 1:U \times 1:N\}$ as $\mathcal{I} \rightarrow \infty$. The limiting variances and covariances are uniformly bounded, using Assumptions A2 and A3. Assumption A1 provides a uniform bound on the discrepancy between $\ell_{u,n}$ and mean of the Gaussian limit. This is enough to derive (4). Assumption A4 gives a stronger bound on covariances between sufficiently distant units, leading to (5). \square

4 ABF-IR theory

Since ABF is ABF-IR with $S = 1$, we focus attention on ABF-IR. At a conceptual level, the localized likelihood estimate in ABF-IR has the same structure as its UBF counterpart. However, ABF-IR additionally requires the capability to satisfactorily implement adapted simulation. Adapted simulation is a local calculation, making it an easier task than the global operation of filtering. Nevertheless, adapted simulation via importance sampling is vulnerable to COD for sufficiently large values of U . For a continuous time model, the use of $S > 1$ is motivated by a result that guided intermediate resampling can reduce, or even remove, the COD in the context of a particle filtering algorithm Park and Ionides (2020). Assumptions B1–B4 are analogous to A1–A4 and are non-asymptotic assumptions

involving $\epsilon_{B1} > 0$, $\epsilon_{B4} > 0$ and $Q > 1$ which are required to hold uniformly over space and time. Assumption B5–B7 control the Monte Carlo error arising from adapted simulation. B5 is a stability property which asserts that the effect of the latent process on the future of the adapted process decays over time. Assumption B6 is a non-asymptotic bound on Monte Carlo error for a single step of adapted simulation. The scaling of the constant \mathcal{C}_0 with U , N and S in Assumption B6 has been studied by Park and Ionides (2020), where it was established that setting $S = U$ can lead to \mathcal{C}_0 being constant, when using an ideal guide function, or slowly growing with U otherwise. The ϵ_{B6}^{-3} error rate in Assumption B6 follows from balancing the two sources of error defined in the statement of Theorem 2 of Park and Ionides (2020). Assumption B7 can be guaranteed by the construction of the algorithm, if independently generated Monte Carlo random variables are used for building the guide function and the one-step prediction particles. The asymptotic limit in Theorem ?? arises as the number of replicates increases.

Assumption B1. *There is an $\epsilon_{B1} > 0$, independent of U and N , and a collection of neighborhoods $\{B_{u,n} \subset A_{u,n}, u \in 1:U, n \in 1:N\}$ such that the following holds for all u and n , and any bounded real-valued function $|h(x)| \leq 1$: if we write $A = A_{u,n}$, $B = B_{u,n}$, $f_A(x_A) = f_{Y_A|X_A}(y_A^*|x_A)$, and $f_B(x_B) = f_{Y_B|X_B}(y_B^*|x_B)$,*

$$\left| \int h(x) \left\{ \frac{\mathbb{E}_g[f_A(X_A^P) f_{X_{u,n}|X_{A[n]}, \mathbf{X}_{n-1}}(x|X_{A[n]}^P, \mathbf{X}_{n-1})]}{\mathbb{E}_g[f_A(X_A^P)]} - \frac{\mathbb{E}_g[f_B(X_B^P) f_{X_{u,n}|X_{B[n]}, \mathbf{X}_{n-1}}(x|X_{B[n]}^P, \mathbf{X}_{n-1})]}{\mathbb{E}_g[f_B(X_B^P)]} \right\} dx \right| < \epsilon_{B1}.$$

Assumption B2. *The bound $\sup_{u \in 1:U, n \in 1:N} |B_{u,n}^+| \leq b$ in Assumption A2 applies for the neighborhoods defined in Assumption B1. This also implies there is a finite maximum temporal depth for the collection of neighborhoods, defined as*

$$d_{\max} = \sup_{(u,n)} \sup_{(\tilde{u}, \tilde{n}) \in B_{u,n}} |n - \tilde{n}|.$$

Assumption B3. *Identically to Assumption A3, $Q^{-1} < f_{Y_{u,n}|X_{u,n}}(y_{u,n}^* | x_{u,n}) < Q$.*

Assumption B4. Define $g_{X_A|X_B}$ via (2). Suppose there is an ϵ_{B4} , independent of U and N , such that the following holds. For each u and n , a set $C_{u,n} \subset (1:U) \times (0:N)$ exists such that $(\tilde{u}, \tilde{n}) \notin C_{u,n}$ implies $B_{u,n}^+ \cap B_{\tilde{u},\tilde{n}}^+ = \emptyset$ and

$$\begin{aligned} |g_{X_{B_{\tilde{u},\tilde{n}}}^P | X_{B_{u,n}}^P} - g_{X_{B_{\tilde{u},\tilde{n}}}^P | X_{B_{u,n}}^P} &< (1/2) \epsilon_{B4} g_{X_{B_{\tilde{u},\tilde{n}}}^P | X_{B_{u,n}}^P} \\ |g_{X_{B_{\tilde{u},\tilde{n}}}^P | \mathbf{X}_{0:N}} g_{X_{B_{u,n}}^P | \mathbf{X}_{0:N}} - g_{X_{B_{\tilde{u},\tilde{n}}}^P, X_{B_{u,n}}^P | \mathbf{X}_{0:N}} & \\ < (1/2) \epsilon_{B4} g_{X_{B_{\tilde{u},\tilde{n}}}^P, X_{B_{u,n}}^P | \mathbf{X}_{0:N}} \end{aligned}$$

Further, there is a uniform bound $|C_{u,n}| \leq c$.

Assumption B5. There is a constant K , independent of U and N , such that, for any $0 \leq d \leq d_{\max}$, any $n \geq K + d$, and any set $D \subset (1:U) \times (n:n-d)$,

$$\begin{aligned} |g_{X_D | \mathbf{X}_{n-d-K}}(x_D | \mathbf{x}_{n-d-K}^{(1)}) - g_{X_D | \mathbf{X}_{n-d-K}}(x_D | \mathbf{x}_{n-d-K}^{(2)})| \\ < \epsilon_{B5} g_{X_D | \mathbf{X}_{n-d-K}}(x_D | \mathbf{x}_{n-d-K}^{(1)}) \end{aligned}$$

holds for all $\mathbf{x}_{n-d-K}^{(1)}$, $\mathbf{x}_{n-d-K}^{(2)}$, and x_D .

Assumption B6. Let h be a bounded function with $|h(x)| \leq 1$. Let $\mathbf{X}_{n,S,j,i}^{\text{IR}}$ be the Monte Carlo quantity constructed in ABF-IR, conditional on $\mathbf{X}_{n-1,S,i}^A = \mathbf{x}_{n-1,S,i}^A$. There is a constant $\mathcal{C}_0(U, N, S)$ such that, for all $\epsilon_{B6} > 0$ and $\mathbf{x}_{n-1,S,i}^A$, whenever the number of particles satisfies $J > \mathcal{C}_0(U, N, S)/\epsilon_{B6}^3$,

$$\left| \mathbb{E} \left[\frac{1}{J} \sum_{j=1}^J h(\mathbf{X}_{n,S,j,i}^{\text{IR}}) \right] - \mathbb{E}_g[h(\mathbf{X}_n) | \mathbf{X}_{n-1} = \mathbf{x}_{n-1,S,i}^A] \right| < \epsilon_{B6}.$$

Assumption B7. For $1 \leq n \leq N$, the Monte Carlo random variable $X_{n,i}^A$ is independent of $w_{u,n,i,j}^M$ conditional on $X_{n-1,i}^A$.

Theorem 2. Let ℓ^{MC} denote the Monte Carlo likelihood approximation constructed by ABF-IR, or by ABF since this is the special case of ABF-IR with $S = 1$. Consider a limit with a growing number of bootstrap replicates, $\mathcal{I} \rightarrow \infty$, and suppose assumptions B1, B2,

B3, B5, B6 and B7. Suppose the number of particles J exceeds the requirement for B6. There are quantities $\epsilon(U, N)$ and $V(U, N)$ with $|\epsilon| < Q^2\epsilon_{B1} + 2Q^{2b}(\epsilon_{B5} + (K + d_{\max})\epsilon_{B6})$ and $V < Q^{4b}U^2N^2$ such that

$$\mathcal{I}^{1/2}[\ell^{MC} - \ell - \epsilon UN] \xrightarrow[\mathcal{I} \rightarrow \infty]{d} \mathcal{N}[0, V].$$

If additionally Assumption B4 holds, we obtain an improved rate of

$$V < Q^{4b}NU\{c + (\epsilon_{B4} + 3\epsilon_{B5} + 4(K + d_{\max})\epsilon_{B6})(NU - c)\}$$

Proof. A full proof is provided in Sec. S4. The extra work to prove Theorem ?? beyond the argument for Theorem 1 is to bound the error arising from the importance sampling approximation to a draw from the adapted transition density. This bound is constructed using Assumptions B5, B6 and B7. The remainder of the proof follows the same approach as Theorem 1, with the adapted process replacing the unconditional latent process. \square

The theoretical results foreshadow our empirical observations that the relative performance of UBF, ABF and ABF-IR is situation-dependent. Assumption A4 is a mixing assumption for the unconditional latent process, whereas Assumption B4 replaces this with a mixing assumption for the adapted process conditional on the data. For a non-stationary process, Assumption A4 may fail to hold uniformly in U whereas the adapted process may provide stable tracking of the latent process (Sec. S2). When Assumption A4 holds, UBF can benefit from not requiring Assumptions B5, B6 and B7. Adapted simulation is an easier problem than filtering, but nevertheless can become difficult in high dimensions, with the consequence that Assumption B6 could require large \mathcal{C}_0 . The tradeoff between ABF and ABF-IR depends on the effectiveness of the guide function for the problem at hand. Intermediate resampling and guide function calculation require additional computational resources, which will necessitate smaller values of \mathcal{I} and J . In some situations, the improved scaling properties of ABF-IR compared to ABF, corresponding to a lower value of \mathcal{C}_0 , will outweigh this cost.

5 Examples

Likelihood evaluation via filtering does not by itself enable parameter estimation for POMP models, however it provides a foundation for Bayesian (Andrieu et al., 2010) and likelihood-based (Ionides et al., 2015) inference. We therefore investigate likelihood evaluation with the expectation that improved methods for likelihood evaluation will lead to improved statistical methodology. Monte Carlo methods for computing the log likelihood generally suffer from both bias and variance, both of which can be considerable for large datasets and complex models. Appropriate inference methodology, such as Monte Carlo adjusted profile (MCAP) confidence intervals, can accommodate substantial Monte Carlo variance so long as the bias is slowly varying across the statistically plausible region of the parameter space (Ionides et al., 2017).

We compare the performance of the three bagged filters (UBF, ABF and ABF-IR) against each other and against alternative approaches. Ensemble Kalman filter (EnKF) methods propagate the ensemble members by simulation from the dynamic model and then update the ensemble to assimilate observations using a Gaussian-inspired rule (Evensen and van Leeuwen, 1996; Lei et al., 2010). The block particle filter (BPF) (Rebeschini and van Handel, 2015; Ng et al., 2002) partitions the latent space and combines independently drawn components from each partition. BPF overcomes COD under weak coupling assumptions (Rebeschini and van Handel, 2015). Unlike these two methods, our bagged filters modify particles only according to the latent dynamics. Thus, our methods satisfy any conservation laws, continuity or smoothness that arise when simulating from the dynamic model.

We first consider a spatiotemporal Gaussian process for which the exact likelihood is available via a Kalman filter. We see in Fig. 1 that ABF-IR can have a considerable advantages over UBF and ABF for problems with an intermediate level of coupling. We then develop a model for measles transmission within and between cities. The model is weakly coupled, leading to successful performance for all three bagged filters. This class of metapopulation models was the primary motivation for the development of these methodologies. However, for a highly coupled system of moderate dimensions, a global particle filter such as GIRF

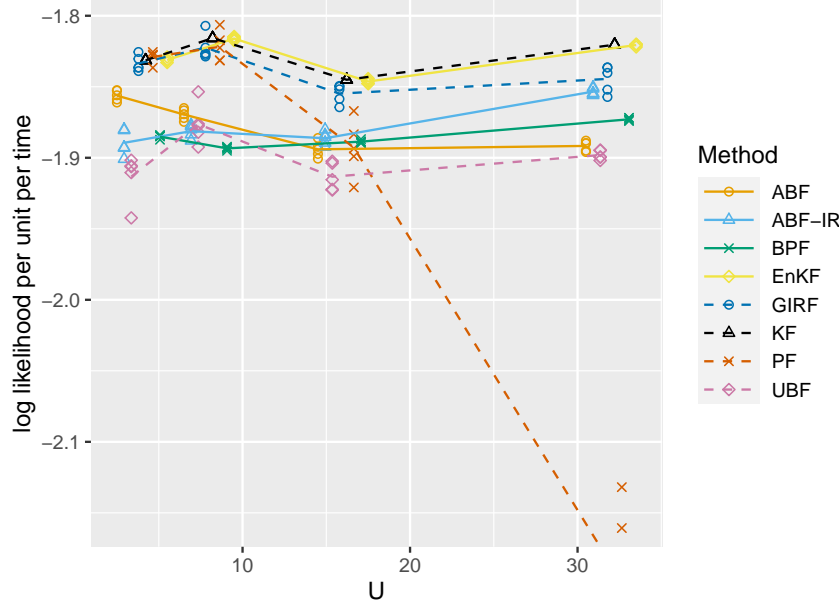


Figure 1: log likelihood estimates for a correlated Brownian motion model of various dimensions. UBF, ABF and ABF-IR are compared with a guided intermediate resampling filter (GIRF), standard particle filter (PF), block particle filter (BPF) and ensemble Kalman filter (EnKF). The exact likelihood was computed via a Kalman filter (KF).

may outperform our bagged methods; an example is the Lorenz-96 model demonstrated in Sec. S8.

5.1 Correlated Brownian motion

Suppose $\mathbf{X}(t) = \Omega \mathbf{W}(t)$ where $\mathbf{W}(t) = W_{1:U}(t)$ comprises U independent standard Brownian motions, and $\Omega_{u,\tilde{u}} = \rho^{d(u,\tilde{u})}$ with $d(u,\tilde{u})$ being the circle distance,

$$d(u, \tilde{u}) = \min(|u - \tilde{u}|, |u - \tilde{u} + U|, |u - \tilde{u} - U|).$$

Set $t_n = n$ for $n = 0, 1, \dots, N$ with initial value $\mathbf{X}(0) = \mathbf{0}$ and suppose measurement errors are independent and normally distributed, $Y_{u,n} = X_{u,n} + \eta_{u,n}$ with $\eta_{u,n} \sim \mathcal{N}(0, \tau^2)$. The parameter ρ determines the strength of the spatial coupling.

Fig. 1 shows how the bagged filters scale on this Gaussian model, compared to a standard particle filter (PF), a guided intermediate resampling filter (GIRF), a block particle filter (BPF), and an ensemble Kalman filter. For our numerical results, we use $\tau = 1$, $\rho = 0.4$ and $N = 20$. The algorithmic parameters and run times are listed in Sec. S5, together with a plot of the simulated data and supplementary discussion. In this case, the exact likelihood is computable via the Kalman filter (KF). Since EnKF is based on a Gaussian approximation, it is also exact in this case, up to a small Monte Carlo error. The GIRF framework encompasses lookahead particle filter techniques, such as the auxiliary particle filter (Pitt and Shepard, 1999), and intermediate resampling techniques (Del Moral et al., 2017). GIRF methods combining these techniques were found to perform better than either of these component techniques alone (Park and Ionides, 2020). Thus, GIRF here represents a state-of-the-art auxiliary particle filter that targets the complete joint filter density for all units. We use the general-purpose, plug-and-play implementation of GIRF provided by the `spatPomp` R package (Asfaw et al., 2019); this implementation of GIRF is thus not tailored in any way to exploit special features of this toy model. PF works well for small values of U in Fig. 1, but rapidly starts struggling as U increases. GIRF behaves comparably to PF for small U but its performance is maintained for larger U . ABF and ABF-IR have some efficiency loss, for small U , relative to PF and GIRF due to the localization involved in the filter weighting, but for large U this cost is paid back by the benefit of the reduced Monte Carlo variability. UBF has a larger efficiency loss for small U , but its favorable scaling properties lead it to overtake ABF for larger U . BPF shows stable scaling and modest efficiency loss. This linear Gaussian SpatPOMP model provides a simple scenario to demonstrate scaling behavior. For filters that cannot take direct advantage of the Gaussian property of the model, we see that there is a tradeoff between efficiency at low U and scalability. This is unavoidable, since there is no known algorithm that is simultaneously fully efficient (up to Monte Carlo error), scalable, and applicable to general SpatPOMP models. We now explore this tradeoff empirically on to a more complex SpatPOMP exemplifying the nonlinear non-Gaussian models motivating our new filtering approach.

5.2 Spatiotemporal measles epidemics

Data analysis for spatiotemporal systems featuring nonlinear, nonstationary mechanisms and partial observability has been a longstanding open challenge for ecological and epidemiological analysis (Bjørnstad and Grenfell, 2001). A compartment modeling framework for spatiotemporal population dynamics divides the population at each spatial location into categories, called compartments, which are modeled as homogeneous. Spatiotemporal compartment models can be called patch models or metapopulation models in an ecological context. Ensemble Kalman filter (EnKF) methods provide a state-of-the-art approach to inference for metapopulation models (Li et al., 2020) despite concerns that the approximations inherent in the EnKF can be problematic for models that are highly nonlinear or non-Gaussian (Ades and Van Leeuwen, 2015). Our bagged filter methodologies have theoretical guarantees for arbitrarily nonlinear and non-Gaussian models, while having improved scaling properties compared to particle filters.

We consider a spatiotemporal model for disease transmission dynamics of measles within and between multiple cities, based on the model of Park and Ionides (2020) which adds spatial interaction to the compartment model presented by He et al. (2010). The model compartmentalizes the population of each city into susceptible (S), exposed (E), infectious (I), and recovered/removed (R) categories. The number of individuals in each compartment city u at time t are denoted by integer-valued random variables $S_u(t)$, $E_u(t)$, $I_u(t)$, and $R_u(t)$. The population dynamics are written in terms of counting processes $N_{\bullet\bullet,u}(t)$ enumerating cumulative transitions in city u , up to time t , between compartments identified by the subscripts. Our model is described by the following system of stochastic differential equations, for $u = 1, \dots, U$,

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

Here, $N_{BS,u}(t)$ models recruitment into the susceptible population, and $N_{\bullet D,u}(t)$ models emigration and death. 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 city u is therefore defined implicitly. $N_{SE,u}(t)$ is modeled as negative binomial death processes (Bretó et al., 2009; Bretó and Ionides, 2011) with over-dispersion parameter σ_{SE} , and rate given by

$$\begin{aligned} \mathbb{E}[N_{SE,u}(t+dt) - N_{SE,u}(t)] &= \beta(t) S_u(t) \left[\left(\frac{I_u + \iota}{P_u} \right)^\alpha \right. \\ &\quad \left. + \sum_{\tilde{u} \neq u} \frac{v_{u\tilde{u}}}{P_u} \left\{ \left(\frac{I_{\tilde{u}}}{P_{\tilde{u}}} \right)^\alpha - \left(\frac{I_u}{P_u} \right)^\alpha \right\} \right] dt + o(dt), \end{aligned} \quad (6)$$

where $\beta(t)$ models seasonality driven by high contact rates between children at school, described by

$$\beta(t) = \begin{cases} (1 + a(1-p)p^{-1}) \bar{\beta} & \text{during school term,} \\ (1-a) \bar{\beta} & \text{during vacation} \end{cases}$$

with $p = 0.759$ being the proportion of the year taken up by the school terms, $\bar{\beta}$ is the mean transmission rate, and a measures the reduction of transmission during school holidays. In (6), α is a mixing exponent modeling inhomogeneous contact rates within a city, and ι models immigration of infected individuals which is appropriate when analyzing a subset of cities that cannot be treated as a closed system. The number of travelers from city u to \tilde{u} is denoted by $v_{u\tilde{u}}$. Here, $v_{u\tilde{u}}$ is constructed using the gravity model of Xia et al. (2004),

$$v_{u\tilde{u}} = G \cdot \frac{\bar{d}}{\bar{P}^2} \cdot \frac{P_u \cdot P_{\tilde{u}}}{d(u, \tilde{u})},$$

where $d(u, \tilde{u})$ denotes the distance between city u and city \tilde{u} , P_u is the average population for city u across time, \bar{P} is the average population across cities, and \bar{d} is the average distance between a randomly chosen pair of cities. Here, we model $v_{u\tilde{u}}$ as fixed through time and symmetric between any two arbitrary cities, though a natural extension would allow for temporal variation and asymmetric movement between two cities. The transition processes $N_{EI,u}(t)$, $N_{IR,u}(t)$ and $N_{\bullet D,u}(t)$ are modeled as conditional Poisson processes with per-capita rates μ_{EI} , μ_{IR} and $\mu_{\bullet D}$ respectively, and we fix $\mu_{\bullet D} = 50 \text{ year}^{-1}$. The birth process $N_{BS,u}(t)$ is an inhomogeneous Poisson processes with rate $\mu_{BS,u}(t)$, given by interpolated census data.

To complete the model specification, we must 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 n th 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

$$\begin{aligned} \mathbb{P}[Y_{u,n}=y \mid Z_{u,n}=z] &= \Phi(y + 0.5; \rho z, \rho(1 - \rho)z + \psi^2 \rho^2 z^2) \\ &\quad - \Phi(y - 0.5; \rho z, \rho(1 - \rho)z + \psi^2 \rho^2 z^2) \end{aligned} \quad (7)$$

where $\Phi(\cdot; \mu, \sigma^2)$ is the $\mathcal{N}(\mu, \sigma^2)$ cumulative distribution function, and ψ models overdispersion relative to the binomial distribution. For $y = 0$, we replace $y - 0.5$ by $-\infty$ in (7).

This model includes many features that have been proposed to be relevant for understanding measles transmission dynamics (He et al., 2010). Our plug-and-play methodology permits consideration of all these features, and readily extends to the investigation of further variations. Likelihood-based inference via plug-and-play methodology therefore provides a framework for evaluating which features of a dynamical model are critical for explaining the data (King et al., 2008). By contrast, Xia et al. (2004) developed a linearization for a specific spatiotemporal measles model which is numerically convenient but not readily adaptable to assess alternative model choices. Fig. 2 shows a simulation from our model, showing that trajectories from this model can capture some features of the system that have been hard to understand: how can it be that disease transmission dynamics between locations have important levels of interaction yet are not locked in synchrony (Becker et al., 2020)? Here, we are developing statistical tools rather than engaging directly in the scientific debate.

We first assess the scaling properties of the filters on the measles model by evaluating the likelihood over varying numbers of units, U , for fixed parameters. The results are given in Fig. 3, with additional information about timing, algorithmic choices, parameter values and a plot of the data provided in Sec. S6. In Fig. 3, the log likelihood per unit per time increases with U because city size decreases with U . Smaller cities have fewer measles cases,

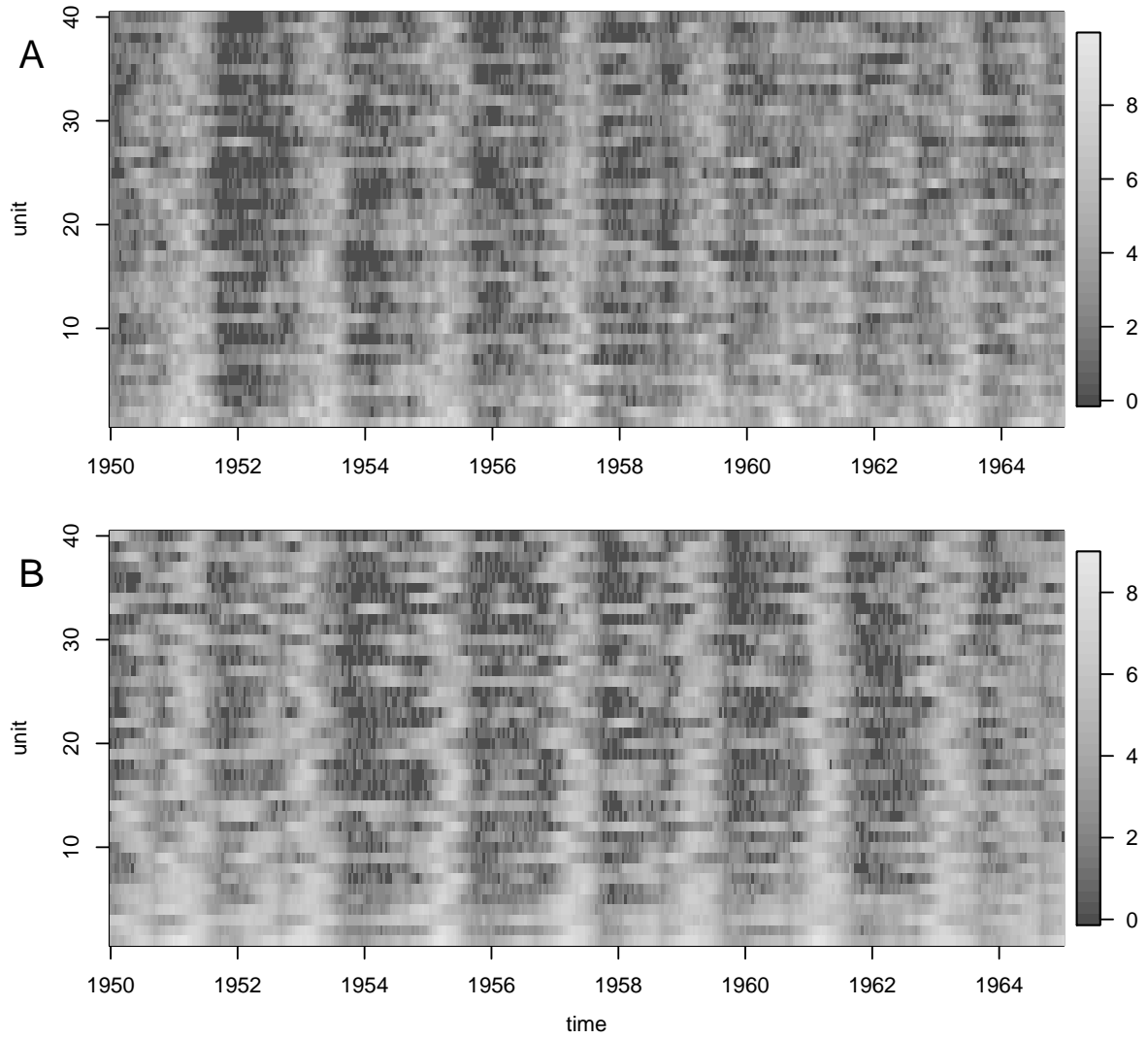


Figure 2: $\text{Log}(\text{reported cases} + 1)$ for (A) the measles simulation used for the likelihood slice; (B) the corresponding UK measles data. The simulation shares the biennial pattern, with most but not all cities locked in phase most of the time.

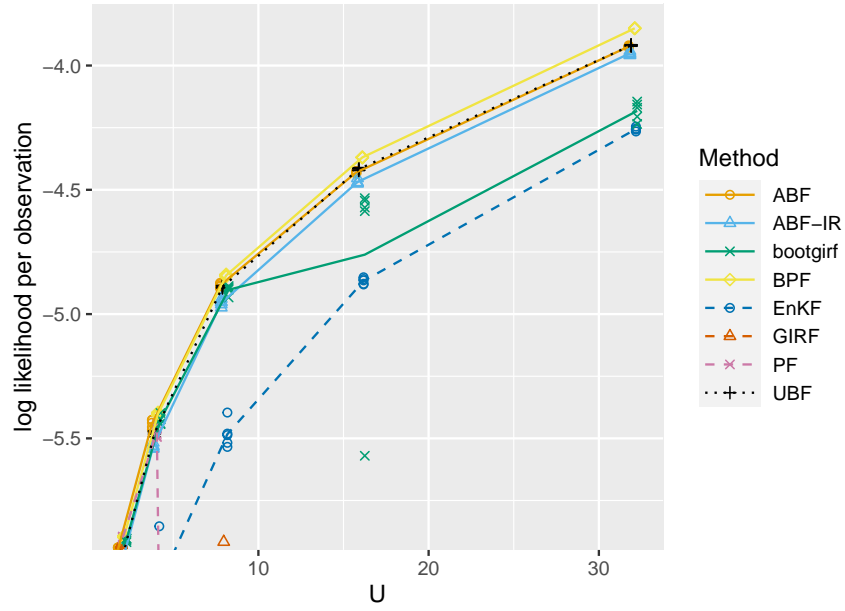


Figure 3: log likelihood estimates for simulated data from the measles model of various dimensions. UBF, ABF and ABF-IR are compared with a guided intermediate resampling filter (GIRF), a standard particle filter (PF), a block particle filter (BPF) and an ensemble Kalman filter (EnKF).

resulting in a narrower and taller probability density function. Fig. 3 shows a rapid decline in the performance of the particle filter (PF) beyond $U = 4$. This is a challenging filtering problem, with dynamics including local fadeouts and high stochasticity in each city stabilized at the metapopulation level by the coupling. In this example, GIRF performs poorly suggesting that the simulated moment guide function is less than successful. We used the general-purpose implementation of GIRF in the `spatPomp` package, and there might be room for improvement by developing a model-specific guide function. ABF-IR uses the same guide function, and this may explain why ABF-IR performs worse than ABF here, though ABF-IR is much less sensitive than GIRF to the quality of the guide. ABF and UBF are competing with BPF as winners on this challenge. The bagged filters and BPF have substantial advantages compared to EnKF, amounting to more than 0.2 log likelihood units per observation. We suspect that the limitations of EnKF on this problem are due to the nonlinearity, non-Gaussianity, and discreteness of fadeout and reintroduction dynamics. All the algorithms have various tuning parameters that could influence the results. Some investigations of alternatives are presented in the Sec. S6. Generalizable conclusions are hard to infer from numerical comparisons of complex algorithms on complex models. Experimentation with different methods, and their tuning parameters, is recommended when investigating a new model.

Fig. 4(A) demonstrates an application of ABF to the task of computing a slice of the likelihood function over the coupling parameter, G , for simulated data. This slice varies G while fixing the other parameters at the values used for the simulation. Scientifically, this gives us an upper bound on the identifiability of G from such data, since the likelihood slice provides statistically efficient inference when all other parameters are known. In this case, the Monte Carlo error is comparable to the change in the log likelihood over a wide range of values of G . Fig. 4(B) shows a similar plot calculated using BPF with comparable computational effort. Both ABF and BPF are successful here, though BPF is more computationally efficient. By contrast, Fig. 4(C) shows that EnKF has substantial bias in estimating G , as well as considerably lower likelihood.

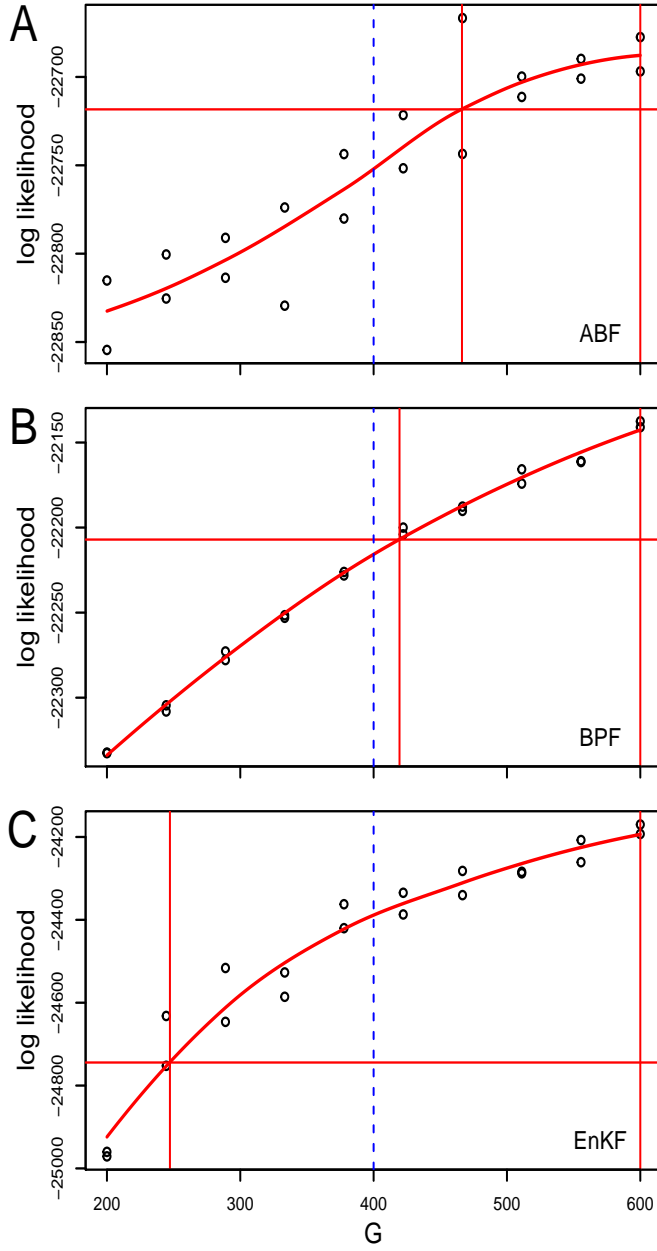


Figure 4: Likelihood slices varying the coupling parameter, for the measles model with $U = 40$ cities, computed via (A) ABF; (B) BPF; (C) EnKF. The solid perpendicular lines construct 95% Monte Carlo adjusted confidence intervals (Ionides et al., 2017). The true parameter value is identified by a blue dashed line.

6 Discussion

The pseudocode presented for the bagged filters describes how the outputs are calculated given the inputs, but does not prescribe details of how these quantities are calculated. There is scope for implementations to trade off memory, computation and communication by making differing decisions on how the various loops defined in the pseudocode are coded, such as decisions on memory over-writing and parallelization. This article focuses on the logical structure of the algorithms, leaving room for future research on implementation-specific considerations, though some supplementary discussion of memory-efficient implementation is given in Sec. S9.

This article has not exhausted the possibilities for bagging different PF algorithms. Bagging a standard PF is one possibility, but the algorithm becomes rapidly degenerate due to COD. Our approach has been to bag PF variants that are already degenerate, so there is less scope for COD to damage them. Alternatively, one could bag PF variants that avoid the COD by using local weights themselves.

We have focused on likelihood evaluation, a critical quantity for testing scientific hypotheses about model structure and the values of unknown parameters. However, our bagged filters also provide localized solutions to the filtering problem, using the local weights to combine estimates across replicates. A solution to the filtering problem, in turn, provides a foundation for forecasting. Effective filtering and likelihood evaluation can enable likelihood maximization via iterated filtering algorithms (Ionides et al., 2011, 2015).

Plug-and-play inference based on sequential Monte Carlo likelihood evaluation has proved successful for investigating highly nonlinear partially observed dynamic systems of low dimension arising in analysis of epidemiological and ecological time series data (Bretó, 2018; Pons-Salort and Grassly, 2018; de Cellès et al., 2018; Marino et al., 2019). A benefit of the plug-and-play property is that it facilitates development of broadly applicable software, which in turn promotes creative scientific model development (Bretó et al., 2009; He et al., 2010).

Geophysical data assimilation has similarities to inference on spatiotemporal biological

systems. Relative to biological systems, geophysical applications are characterized by a greater number of spatial locations, better mathematical understanding of the underlying processes, and lower stochasticity. From this literature, the locally weighted particle filter of (Poterjoy, 2016) is perhaps closest to our approach, but those local weights are used to construct a localized Kalman gain which is motivated by a Gaussian approximation.

The EnKF algorithm arose originally via geophysical research (Evensen and van Leeuwen, 1996) and has since become used more widely for inference on SpatPOMP models (Katzfuss et al., 2020; Lei et al., 2010). However, EnKF can fail entirely even on simple POMP models if the structure is sufficiently non-Gaussian. For example, let X_n be a one-dimensional Gaussian random walk, and let Y_n given $X_n = x_n$ be normally distributed with mean 0 and variance x_n^2 . The linear filter rule used by EnKF to update the estimate of X_n given Y_n has mean zero for any value of X_n , since X_n and Y_n are uncorrelated. Therefore, the EnKF filter estimate of the latent process remains essentially constant regardless of the data. Models of this form are used in finance to describe stochastic volatility. EnKF could be applied more successfully by modifying model, such as replacing Y_n by $|Y_n|$, but for complex models it may be unclear whether and where such problems are arising. Our results show that there is room for improvement over EnKF on a spatiotemporal epidemiology model, though in our example there is no clear advantage for BF methods over BPF.

Latent state trajectories constructed in our BF algorithms are all generated from the model simulator, appropriately reweighted and resampled, and so are necessarily valid sample paths of the model. For example, spatial smoothness properties of the model through space, or conservation properties where some function of the system remains unchanged through time, are maintained in the BF trajectories. This is not generally true for the block particle filter (since resampling blocks can lead to violations at block boundaries) or for EnKF (since the filter procedure perturbs particles using a linear update rule that cannot respect nonlinear relationships). The practical importance of smoothness and conservation considerations will vary with the system under investigation, but this property of BF gives the scientific investigator one less thing to worry about.

The algorithms UBF, ABF, ABF-IR, GIRF, PF, BPF, and EnKF compared in this article all enjoy the plug-and-play property, facilitating their implementations in general-purpose software. The numerical results for this paper use the `abf`, `abfir`, `girf`, `pfilter`, `bpfilter` and `enkf` functions via the open-source R package `spatPomp` (Asfaw et al., 2019) that provides a spatiotemporal extension of the R package `pomp` (King et al., 2016). UBF was implemented using `abf` with $J = 1$ particles per replicate. The source code for this paper will be contributed to an open-source scientific archive upon acceptance for publication.

References

- Ades, M. and Van Leeuwen, P. J. (2015). The equivalent-weights particle filter in a high-dimensional system. *Quarterly Journal of the Royal Meteorological Society*, 141(687):484–503.
- Andrieu, C., Doucet, A., and Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 72:269–342.
- Asfaw, K., Ionides, E. L., and King, A. A. (2019). `spatPomp`: R package for statistical inference for spatiotemporal partially observed Markov processes. <https://github.com/kidusasfaw/spatPomp>.
- 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.

- Bjørnstad, O. N. and Grenfell, B. T. (2001). Noisy clockwork: Time series analysis of population fluctuations in animals. *Science*, 293:638–643.
- Brehmer, J., Louppe, G., Pavez, J., and Cranmer, K. (2020). Mining gold from implicit models to improve likelihood-free inference. *Proceedings of the National Academy of Sciences of the USA*, 117(10):5242–5249.
- Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2):123–140.
- Bretó, C. (2018). Modeling and inference for infectious disease dynamics: A likelihood-based approach. *Statistical Science*, 33(1):57–69.
- 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.
- Chandler, R. E. (2013). Exploiting strength, discounting weakness: Combining information from multiple climate simulators. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1991):20120388.
- Crisan, D. and Rozovskii, B. (2011). *The Oxford handbook of nonlinear filtering*. Oxford University Press.
- de Cellès, M. D., Magpantay, F. M., King, A. A., and Rohani, P. (2018). The impact of past vaccination coverage and immunity on pertussis resurgence. *Science Translational Medicine*, 10(434):eaaj1748.
- Del Moral, P., Moulines, E., Olsson, J., and Vergé, C. (2017). Convergence properties of weighted particle islands with application to the double bootstrap algorithm. *Stochastic Systems*, 6(2):367–419.

- Del Moral, P. and Murray, L. M. (2015). Sequential Monte Carlo with highly informative observations. *Journal on Uncertainty Quantification*, 3:969–997.
- Doucet, A., de Freitas, N., and Gordon, N. J. (2001). *Sequential Monte Carlo Methods in Practice*. Springer, New York.
- 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.
- Ebert, E. E. (2001). Ability of a poor man’s ensemble to predict the probability and distribution of precipitation. *Monthly Weather Review*, 129(10):2461–2480.
- Evensen, G. and van Leeuwen, P. J. (1996). Assimilation of geostat altimeter data for the Agulhas Current using the ensemble Kalman filter with a quasigeostrophic model. *Monthly Weather Review*, 124:58–96.
- Gordon, N., Salmond, D. J., and Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEEE Proceedings-F*, 140(2):107–113.
- 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., Bhadra, A., Atchadé, Y., and King, A. A. (2011). Iterated filtering. *Annals of Statistics*, 39:1776–1802.
- 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.

- 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.
- Kevrekidis, I. G. and Samaey, G. (2009). Equation-free multiscale computation: Algorithms and applications. *Annual Review of Physical Chemistry*, 60:321–344.
- King, A. A., Ionides, E. L., Pascual, M., and Bouma, M. J. (2008). Inapparent infections and cholera dynamics. *Nature*, 454:877–880.
- 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.
- Lei, J., Bickel, P., and Snyder, C. (2010). Comparison of ensemble Kalman filters under non-Gaussianity. *Monthly Weather Review*, 138(4):1293–1306.
- Leutbecher, M. and Palmer, T. N. (2008). Ensemble forecasting. *Journal of Computational Physics*, 227(7):3515–3539.
- 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.
- Marino, J. A., Peacor, S. D., Bunnell, D. B., Vanderploeg, H. A., Pothoven, S. A., Elgin, A. K., Bence, J. R., Jiao, J., and Ionides, E. L. (2019). Evaluating consumptive and nonconsumptive predator effects on prey density using field times series data. *Ecology*, 100(3):e02583.
- Ng, B., Peshkin, L., and Pfeffer, A. (2002). Factored particles for scalable monitoring. *Proceedings of the 18th Conference on Uncertainty and Artificial Intelligence*, pages 370–377.

- Palmer, T. N. (2002). The economic value of ensemble forecasts as a tool for risk assessment: From days to decades. *Quarterly Journal of the Royal Meteorological Society*, 128(581):747–774.
- 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.
- Pitt, M. K. and Shepard, N. (1999). Filtering via simulation: Auxillary particle filters. *Journal of the American Statistical Association*, 94:590–599.
- Pons-Salort, M. and Grassly, N. C. (2018). Serotype-specific immunity explains the incidence of diseases caused by human enteroviruses. *Science*, 361(6404):800–803.
- Poterjoy, J. (2016). A localized particle filter for high-dimensional nonlinear systems. *Monthly Weather Review*, 144(1):59–76.
- 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.
- 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.