Bagged filters for partially observed interacting systems

Edward L. Ionides*
Department of Statistics, University of Michigan
Kidus Asfaw
Department of Statistics, University of Michigan
Joonha Park
Department of Mathematics, University of Kansas
Aaron A. King
Department of Ecology and Evolutionary Biology &
Center for the Study of Complex Systems, University of Michigan

Abstract

Bagging (i.e., bootstrap aggregating) involves combining an ensemble of bootstrap estimators. We consider bagging for inference from noisy or incomplete measurements on a collection of interacting stochastic dynamic systems. Each system is called a unit, and each unit is associated with a spatial location. A motivating example arises in epidemiology, where each unit is a city: the majority of transmission occurs within a city, with smaller yet epidemiologically important interactions arising from disease transmission between cities. Monte Carlo filtering methods used for inference on nonlinear non-Gaussian systems can suffer from a curse of dimensionality as the number of units increases. We introduce bagged filter (BF) methodology which combines an ensemble of Monte Carlo filters, using spatiotemporally localized weights to select successful filters at each unit and time. We obtain conditions under which likelihood evaluation using a BF algorithm can beat a curse of dimensionality, and we demonstrate applicability even when these conditions do not hold. BF can out-perform an ensemble Kalman filter on a coupled population dynamics model describing infectious disease transmission. A block particle filter also 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, 1-U01-GM110712 and 1-R01-AI143852. We acknowledge constructive feedback from two anonymous referees and the associate editor.

1 Introduction

Bagging is a technique to improve numerically unstable estimators by combining an ensemble of replicated bootstrap calculations (Breiman, 1996). In the context of nonlinear partially observed dynamic systems, the bootstrap filter of Gordon et al. (1993) has led to a variety of particle filter (PF) methodologies (Doucet et al., 2001; Doucet and Johansen, 2011); Here, we consider algorithms combining an ensemble of replicated particle filters, which we term *bagged filter* algorithms. Standard PF methods suffer from a curse of dimensionality (COD), defined as an exponential increase in computational requirement as the problem size grows, limiting its applicability to large 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 affordable numbers of particles. 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). Similar properties to plug-and-play are *likelihood-free* (Brehmer et al., 2020) and *equation-free* (Kevrekidis and Samaey, 2009). The ensemble Kalman filter (Evensen, 2009; Lei et al., 2010; Katzfuss et al., 2020) is a widely used plug-and-play method which uses simulations to construct a nonlinear filter that is exact for a linear Gaussian model. 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, making localization approximations only when comparing these trajectories to data.

The replicated stochastic trajectories in a bagged filter form an ensemble of representations of the dynamic system. Unlike the particles in a particle filter or ensemble Kalman filter, 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 collec-

tion 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 provide effective filtering 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.

We first consider a simple bagged filter where each replicate is an independent simulation of the latent process model. We call this the unadapted bagged filter (UBF) since the replicates in the ensemble depend on the model but not on the data. UBF is described in Sec. 2, with a theoretical analysis presented in Sec. 2.1. Each UBF replicate corresponds to a basic PF algorithm with a single particle. We show that UBF formally beats the COD under a weak mixing assumption, though UBF can have poor numerical behavior if a very large number of replicates are needed to reach this happy asymptotic limit. Subsequent empirical results show that UBF may nevertheless be a useful algorithm in some situations. In Sec. 3, we generalize UBF to construct an adapted bagged filter (ABF) where each replicate tracks the data. The price of adaptation is that ABF no longer avoids the COD, a limitation that can be controlled in certain situations by supplementing ABF with a technique called intermediate resampling, to obtain the ABF-IR algorithm. Theoretical results for ABF and ABF-IR algorithms are developed in Sec. 3.1. The algorithms are demonstrated in action and compared with alternative approaches in Sec. 4.

2 The unadapted bagged filter (UBF)

Suppose the collection of units is indexed by the set $\{1, 2, \ldots, U\}$, which is written as $1\!:\!U$. The 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 $\mathbb{X}^U$. This discrete time process may arise from a continuous time Markov process $\{\boldsymbol{X}(t), t_0 \le t \le t_N\}$ observed at times $t_{1:N}$, and in this 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 $\{\boldsymbol{Y}_n = Y_{1:U,n}, n \in 1:N\}$ which takes values in a product space $\mathbb{Y}^U$. Observations are modeled as being conditionally independent given the latent process. The conditional independence of measurements applies over both time and the unit structure, so the collection $\big\{Y_{u,n}, u \in 1:U, n \in 1:N\big\}$ is conditionally independent given $\big\{X_{u,n}, u \in 1:U, n \in 1:N\big\}$. The unit structure for the observation process is not necessary for all that follows (see Sec. S1). We suppose the existence of a joint density $f_{\boldsymbol{X}_{0:N},\boldsymbol{Y}_{1:N}}$ of $X_{1:U,0:N}$ 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. 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, Bretó et al., 2009), 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, and so we call the model a SpatPOMP. In the following, we use a lexicographical ordering on the set of observations; Specifically, we define the set of observations preceding unit u at time n as

$$A_{u,n} = \big\{(\tilde{u}, \tilde{n}) : 1 \le \tilde{n} < n \text{ or } (\tilde{n} = n \text{ and } \tilde{u} < u)\big\}. \tag{1}$$

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. The bagged filter algorithms we study do not evaluate or simulate such transition densities but only compute the measurement model on neighborhoods, unlike the filter of Beskos et al. (2017) built on a similar factorization. If sufficiently distant units are approximately independent, we say the system is *weakly coupled*. In this case, 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 $B_{u,n}$.

Our primary interest is estimation of the log likelihood for the data given the model, $\ell = \log f_{\boldsymbol{Y}_{1:N}}(\boldsymbol{y}^*_{1:N})$, which is of fundamental importance in both Bayesian and non-Bayesian statistical inference. A general filtering problem is to evaluate $\mathbb{E}\big[h(X_{u,n}) \,|\, Y_{A_{u,n}} = y^*_{A_{u,n}}\big]$ for

some function $h : \mathbb{X} \to \mathbb{R}$. Taking $h(x) = f_{Y_{u,n}|X_{u,n}}\left(y^*_{u,n} \,|\, x\right)$ gives a filtering representation of the likelihood evaluation problem. Further discussion on bagged filtering for other filtering problems is given in Sec. S10. For likelihood-based inference, maximization plays an important role in point estimation, confidence interval construction, hypothesis testing and model selection. An extension of bagged filtering to likelihood maximization is demonstrated in Sec. 4.3 following the approach described in Sec. S11.

Pseudocode for a UBF algorithm for likelihood evaluation is given below. The prediction weight $w^P_{u,n,i}$ gives an appropriate weighting for replicate i for predicting $y^*_{u,n}$ based on the most relevant data, $y^*_{B_{u,n}}$. Conditional log likelihoods are estimated using an approximation

$$\begin{aligned}
\ell_{u,n} &= \log f_{Y_{u,n}|Y_{A_{u,n}}}\left(y^*_{u,n} \,|\, y^*_{A_{u,n}}\right) = \log\left(\int f_{Y_{u,n}|X_{u,n}}(y^*_{u,n} \,|\, x)\, f_{X_{u,n}|Y_{A_{u,n}}}(x \,|\, y^*_{A_{u,n}})\, dx\right) \\
&\approx \log\left(\int f_{Y_{u,n}|X_{u,n}}(y^*_{u,n} \,|\, x)\, f_{X_{u,n}|Y_{B_{u,n}}}(x \,|\, y^*_{B_{u,n}})\, dx\right).
\end{aligned}$$

The choice of $B_{u,n}$ is determined empirically, with a bias-variance trade-off used to compare small neighborhoods such as $B_{u,n} = \{(u, n-1), (u-1, n)\}$ or $B_{u,n} = \{(u, n-1), (u, n-2)\}$ against larger neighborhoods. The plug-and-play property is evident because UBF requires as input a simulator for the latent coupled dynamic process but not an evaluator of transition probabilities.

Algorithm 1: Unadapted bagged filter (UBF).

input: simulator for $f_{\boldsymbol{X}_n|\boldsymbol{X}_{n-1}}(\boldsymbol{x}_n \,|\, \boldsymbol{x}_{n-1})$ and $f_{\boldsymbol{X}_0}(\boldsymbol{x}_0)$; evaluator for $f_{Y_{u,n}|X_{u,n}}(y_{u,n} \,|\, x_{u,n})$; data, $\boldsymbol{y}^*_{1:N}$; number of replicates, $\mathcal{I}$; neighborhood structure, $B_{u,n}$

for i in $1{:}\mathcal{I}$ **do**

  initialize simulation, $\boldsymbol{X}_{0,i} \sim f_{\boldsymbol{X}_0}(\cdot)$

  for n in $1{:}N$ **do**

    simulate, $\boldsymbol{X}_{n,i} \sim f_{\boldsymbol{X}_n|\boldsymbol{X}_{n-1}}\big(\cdot \,|\, \boldsymbol{X}_{n-1,i}\big)$

    measurement weights, $w^M_{u,\tilde{n},i} = f_{Y_{u,n}|X_{u,n}}\big(y^*_{u,n} \,|\, X_{u,n,i}\big)$ for u in $1{:}U$

    prediction weights, $w^P_{u,n,i} = \prod_{(\tilde{u},\tilde{n})\in B_{u,n}} w^M_{\tilde{u},\tilde{n},i}$ for u in $1{:}U$

  end

end

$\ell^{\mathrm{MC}}_{u,n} = \log\Big(\sum_{i=1}^{\mathcal{I}} w^M_{u,n,i} w^P_{u,n,i}\Big) - \log\Big(\sum_{i=1}^{\mathcal{I}} w^P_{u,n,i}\Big)$ for u in $1{:}U$, n in $1{:}N$

output: log likelihood estimate, $\ell^{\mathrm{MC}} = \sum_{n=1}^{N}\sum_{u=1}^{U} \ell^{\mathrm{MC}}_{u,n}$

2.1 UBF theory

A dataset $\boldsymbol{y}^*_{1:N}$ with U units is modeled via a joint density $f_{\boldsymbol{X}_{0:N},\boldsymbol{Y}_{1:N}}$. We consider non-asymptotic bounds that apply for all values of U and N. To impose a requirement that distant regions of space-time behave similarly and have only weak dependence, we assert the following conditions which define constants ϵ_{A1}, ϵ_{A4} and Q used to bound the 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 $\epsilon_{\mathrm{A1}} > 0$, 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* $x_{B^c_{u,n}}$*,*

$$\left| \int h(x_{u,n}) f_{X_{u,n}|Y_{B_{u,n}},X_{B^c_{u,n}}}(x_{u,n} \,|\, y^*_{B_{u,n}}, x_{B^c_{u,n}})\, dx_{u,n} \right.$$
$$\left. - \int h(x_{u,n}) f_{X_{u,n}|Y_{B_{u,n}}}(x_{u,n} \,|\, y^*_{B_{u,n}})\, dx_{u,n} \right| < \epsilon_{\mathrm{A1}}.$$

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*

$$\sup_{u \in 1:U,\, n \in 1:N} \left| B^+_{u,n} \right| \leq b.$$

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* $\epsilon_{\mathrm{A4}} > 0$*, 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*

$$\left| f_{X_{B^+_{\tilde{u},\tilde{n}}}|X_{B^+_{u,n}}} - f_{X_{B^+_{\tilde{u},\tilde{n}}}} \right| < \epsilon_{\mathrm{A4}}\, f_{X_{B^+_{\tilde{u},\tilde{n}}}}$$

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 describes 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, $\mathcal{I} \to \infty$, and suppose assumptions A1, A2 and A3. There are quantities $\epsilon(U,N)$ and $V(U,N)$, with bounds $|\epsilon| < \epsilon_{\mathrm{A1}} Q^2$ and $V < Q^{4b} U^2 N^2$, such that*

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

where $\xrightarrow[\mathcal{I}\to\infty]{d}$ 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

$$V < Q^{4b} UN\big(c + \epsilon_{\mathrm{A4}}\,(UN - c)\big). \tag{3}$$

Proof. A complete proof is given in Sec. S3. Briefly, the assumptions imply a multivariate central limit theorem for $\{\ell^{\mathrm{MC}}_{u,n}, (u,n) \in 1{:}U \times 1{:}N\}$ as $\mathcal{I} \to \infty$. 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 (2). Assumption A4 gives a stronger bound on covariances between sufficiently distant units, leading to (3). □

Theorem 1 does not guarantee uniformity over U and N of the rate of convergence as $\mathcal{I} \to \infty$. However, it does guarantee that the polynomial bounds in (2) and (3) hold for sufficiently large $\mathcal{I}$. The COD is characterized by exponential bounds, and so Theorem 1 shows a specific sense in which UBF can avoid COD. Uniformity of the central limit convergence in Theorem 1 may be expected to hold via a Berry-Esseen theorem, but extension of existing Berry-Esseen results for dependent processes (Bentkus et al., 1997; Jirak, 2016) is beyond the scope of this article.

The approximation error for UBF 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. 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 constraints satisfied by the model for the latent process.

3 Adaptation and intermediate resampling

Theorem 1 shows that UBF can beat COD. However, UBF can perform poorly on long time series unless weak temporal dependence allows simulated sample paths to remain relevant over the course of a long time series. For example, we will find that UBF performs well on an epidemiological model (Sec. 5) but less well on a geophysical model (Sec. S8). It is sometimes 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 of drawing from $f_{\boldsymbol{X}_n|\boldsymbol{Y}_{1:n}}$ may be intractable via importance sampling methods, a version of this problem localized in space and time may nevertheless be feasible. The conditional density, $f_{\boldsymbol{X}_n|\boldsymbol{Y}_n,\boldsymbol{X}_{n-1}}$, is called the *adapted density*, and simulating from this density is called *adapted simulation*. For models where $\boldsymbol{X}_{n-1}$ is highly informative about $\boldsymbol{X}_n$, importance sampling for adapted simulation may be much easier than the full filter calculation. The following adapted bagged filter (ABF) is constructed under a hypothesis that the adapted simulation problem is tractable, and it is applicable when the number of units is prohibitive for Monte Carlo sampling from the full filter distribution but not for sampling from the adapted distribution. In ABF, the adapted simulations are reweighted in a neighborhood of each unit and time point to construct a local approximation to the filtering problem which leads to an estimate of the likelihood. The pseudocode for ABF, below, reduces to UBF when using a single particle per repicate, $J = 1$.

Algorithm 2: Adapted bagged filter (ABF)

input: simulator for $f_{\boldsymbol{X}_n|\boldsymbol{X}_{n-1}}(\boldsymbol{x}_n \,|\, \boldsymbol{x}_{n-1})$ and $f_{\boldsymbol{X}_0}(\boldsymbol{x}_0)$; evaluator for $f_{Y_{u,n}|X_{u,n}}(y_{u,n} \,|\, x_{u,n})$; data, $\boldsymbol{y}^*_{1:N}$; number of particles per replicate, J; number of replicates, $\mathcal{I}$; neighborhood structure, $B_{u,n}$

for i in $1{:}\mathcal{I}$ **do**

  initialize adapted simulation, $\boldsymbol{X}^{\mathrm{A}}_{0,i} \sim f_{\boldsymbol{X}_0}(\cdot)$

  for n in $1{:}N$ **do**

    proposals: $\boldsymbol{X}^{\mathrm{P}}_{n,i,j} \sim f_{\boldsymbol{X}_n|X_{1:U,n-1}}\big(\boldsymbol{x}_n \,|\, \boldsymbol{X}^{\mathrm{A}}_{n-1,i}\big)$ for j in $1{:}J$

    $w^{M}_{u,n,i,j} = f_{Y_{u,n}|X_{u,n}}\big(y^*_{u,n} \,|\, X^{\mathrm{P}}_{u,n,i,j}\big)$ for u in $1{:}U$, j in $1{:}J$

    adapted resampling weights, $w^{\mathrm{A}}_{n,i,j} = \prod_{u=1}^{U} w^{M}_{u,n,i,j}$ for u in $1{:}U$, j in $1{:}J$

    $\boldsymbol{X}^{\mathrm{A}}_{n,i} = \boldsymbol{X}^{\mathrm{P}}_{n,i,r(i)}$ with $\mathbb{P}\big[r(i)=a\big] = w^{\mathrm{A}}_{n,i,a}\Big/\sum_{k=1}^{J} w^{\mathrm{A}}_{n,i,k}$

    $w^{\mathrm{P}}_{u,n,i,j} = \prod_{\tilde{n}=1}^{n-1}\Big[\frac{1}{J}\sum_{k=1}^{J}\prod_{(\tilde{u},\tilde{n})\in B_{u,n}} w^{M}_{\tilde{u},\tilde{n},i,k}\Big]\prod_{(\tilde{u},n)\in B_{u,n}} w^{M}_{\tilde{u},n,i,j}$ for u in $1{:}U$, j in $1{:}J$

  end

end

$\ell^{\mathrm{MC}}_{u,n} = \log\left(\dfrac{\sum_{i=1}^{\mathcal{I}}\sum_{j=1}^{J} w^{M}_{u,n,i,j} w^{P}_{u,n,i,j}}{\sum_{i=1}^{\mathcal{I}}\sum_{j=1}^{J} w^{P}_{u,n,i,j}}\right)$ for u in $1{:}U$, n in $1{:}N$

output: log likelihood estimate, $\ell^{\mathrm{MC}} = \sum_{n=1}^{N}\sum_{u=1}^{U}\ell^{\mathrm{MC}}_{u,n}$

ABF remedies a weakness of UBF by making each boostrap filter adapted to the data. However, this benefit carries a cost, since adapted simulation is not immune from the curse of dimensionality. Therefore, we also consider an algorithm called ABF-IR which uses an intermediate resampling technique to carry out the adapted simulation. 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 process has a continuous time representation, $\{\boldsymbol{X}(t)\}$, with observation times $t_{1:N}$. We write S intermediate resampling times as

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

Carrying out an intermediate resampling procedure can have favorable scaling properties

when S is proportional to 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). Only an ideal guide function fully addresses COD (Park and Ionides, 2020) and on nontrivial problems this is not available. However, practical guide functions can nevertheless improve performance.

The implementation in the ABF-IR pseudocode constructs the guide $g_{n,s,i,j}$ using a simulated moment method proposed by Park and Ionides (2020). The quantities $\boldsymbol{X}^G_{n,i,j}$, $V_{u,n,i}$, $\boldsymbol{\mu}^{\mathrm{IP}}_{n,s,i,j}$, $V^{\mathrm{meas}}_{u,n,s,i,j}$, $V^{\mathrm{proc}}_{u,n,s,i}$ 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

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

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

$$\begin{aligned} h_{u,n}(x_{u,n}) &= \mathbb{E}\big[Y_{u,n} \,|\, X_{u,n} = x_{u,n}\big] \\ \overrightarrow{\mathrm{v}}_{u,n}(x_{u,n}, \theta) &= \mathrm{Var}\big(Y_{u,n} \,|\, X_{u,n} = x_{u,n}\,; \theta\big) \end{aligned}$$

Also required for ABF-IR is an inverse function $\overleftarrow{\mathrm{v}}_{u,n}\big(V, x_{u,n}, \theta\big)$ such that $\overleftarrow{\mathrm{v}}_{u,n}\big(V, x_{u,n}, \theta\big) = \phi$ implies

$$\overrightarrow{\mathrm{v}}_{u,n}\big(x_{u,n}, \phi\big) = V.$$

Algorithm 3: ABF with intermediate resampling (ABF-IR)

input: same as Algorithm 2 (ABF) plus: intermediate timesteps, S; measurement variance functions, $\overleftarrow{\mathrm{v}}_{u,n}$ and $\overrightarrow{\mathrm{v}}_{u,n}$; approximate mean functions, $\boldsymbol{\mu}$ and $h_{u,n}$

for i in $1{:}\mathcal{I}$ **do**

  initialize adapted simulation, $\boldsymbol{X}^{\mathrm{A}}_{0,i} \sim f_{\boldsymbol{X}_0}(\cdot)$

  for n in $1{:}N$ **do**

    guide simulations: $\boldsymbol{X}^{G}_{n,i,j} \sim f_{\boldsymbol{X}_n|\boldsymbol{X}_{n-1}}\big(\boldsymbol{x}_n \,|\, \boldsymbol{X}^{\mathrm{A}}_{n-1,i}\big)$ for j in $1{:}J$

    $V_{u,n,i} = \mathrm{Var}\big\{h_{u,n}\big(X^{G}_{u,n,i,j}\big),\ j \text{ in } 1{:}J\big\}$

    $g^{\mathrm{R}}_{n,0,i,j} = 1 \quad \text{and} \quad \boldsymbol{X}^{\mathrm{IR}}_{n,0,i,j} = \boldsymbol{X}^{\mathrm{A}}_{n-1,i}$ for j in $1{:}J$

    for s in $1{:}S$ **do**

      $\boldsymbol{X}^{\mathrm{IP}}_{n,s,i,j} \sim f_{\boldsymbol{X}_{n,s}|\boldsymbol{X}_{n,s-1}}\big(\,\cdot\,|\boldsymbol{X}^{\mathrm{IR}}_{n,s-1,i,j}\big)$ for j in $1{:}J$

      $\boldsymbol{\mu}^{\mathrm{IP}}_{n,s,i,j} = \boldsymbol{\mu}\big(\boldsymbol{X}^{\mathrm{IP}}_{n,s,i,j}, t_{n,s}, t_n\big)$ for j in $1{:}J$

      $V^{\mathrm{meas}}_{u,n,s,i,j} = \overrightarrow{\mathrm{v}}_u(\theta, \mu^{\mathrm{IP}}_{u,n,s,i,j})$ for u in $1{:}U$, j in $1{:}J$

      $V^{\mathrm{proc}}_{u,n,s,i} = V_{u,n,i}\,\big(t_n - t_{n,s}\big)\Big/\big(t_n - t_{n,0}\big)$ for u in $1{:}U$

      $\theta_{u,n,s,i,j} = \overleftarrow{\mathrm{v}}_u\big(V^{\mathrm{meas}}_{u,n,s,i,j} + V^{\mathrm{proc}}_{u,n,s,i},\, \mu^{\mathrm{IP}}_{u,n,s,i,j}\big)$ for u in $1{:}U$, j in $1{:}J$

      $g_{n,s,i,j} = \prod_{u=1}^{U} f_{Y_{u,n}|X_{u,n}}\big(y^*_{u,n}\,|\,\mu^{\mathrm{IP}}_{u,n,s,i,j}\,;\theta_{u,n,s,i,j}\big)$ for j in $1{:}J$

      guide weights: $w^{G}_{n,s,i,j} = g_{n,s,i,j}/g^{\mathrm{R}}_{n,s-1,i,j}$ for j in $1{:}J$

      resampling: $\mathbb{P}\big[r(i,j)=a\big] = w^{G}_{n,s,i,a}\Big/\Big(\sum_{k=1}^{J} w^{G}_{n,s,i,k}\Big)^{-1}$ for j in $1{:}J$

      $\boldsymbol{X}^{\mathrm{IR}}_{n,s,i,j} = \boldsymbol{X}^{\mathrm{IP}}_{n,s,i,r(i,j)} \quad \text{and} \quad g^{\mathrm{R}}_{n,s,i,j} = g_{n,s,i,r(i,j)}$ for j in $1{:}J$

    end

    $\boldsymbol{X}^{\mathrm{A}}_{n,i} = \boldsymbol{X}^{\mathrm{IR}}_{n,S,i,1}$

    $w^{M}_{u,n,i,j} = f_{Y_{u,n}|X_{u,n}}\big(y^*_{u,n}\,|\,X^{G}_{u,n,i,j}\big)$ for u in $1{:}U$, j in $1{:}J$

    $w^{\mathrm{P}}_{u,n,i,j} = \prod_{\tilde{n}=1}^{n-1}\Big[\frac{1}{J}\sum_{a=1}^{J}\prod_{\tilde{u}:(\tilde{u},\tilde{n})\in B_{u,n}} w^{M}_{\tilde{u},\tilde{n},i,a}\Big]\prod_{\tilde{u}:(\tilde{u},n)\in B_{u,n}} w^{M}_{\tilde{u},n,i,j}$ for u in $1{:}U$, j in $1{:}J$

  end

end

output: $\ell^{\,\mathrm{MC}} = \sum_{u=1}^{U}\sum_{n=1}^{N}\log\left(\frac{\sum_{i=1}^{\mathcal{I}}\sum_{j=1}^{J} w^{M}_{u,n,i,j} w^{P}_{u,n,i,j}}{\sum_{i=1}^{\mathcal{I}}\sum_{j=1}^{J} w^{P}_{u,n,i,j}}\right)$

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).

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. An answer is that adaptive simulation can fail to track the observation sequence when one resamples down to a single particle too often (Sec. S2).

3.1 ABF-IR theory

We start by considering a deterministic limit for infinite Monte Carlo effort and explaining why the 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 $A^+_{u,n} = A_{u,n} \cup (u,n)$, matching the defintion $B^+_{u,n} = B_{u,n} \cup (u,n)$. 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 $B^{[m]} = B \cap \big(1\!:\!U \times \{m\}\big)$. ABF and ABF-IR build on the following identity,

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

where $f_{\boldsymbol{X}_m|\boldsymbol{X}_{m-1},\boldsymbol{Y}_m}$ 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 $g_{\boldsymbol{X}_{0:N},\boldsymbol{X}^P_{1:N}}(\boldsymbol{x}_{0:N}, \boldsymbol{x}^P_{1:N})$ be the joint density of the adapted process and the proposal process,

$$\begin{aligned} g_{\boldsymbol{X}_{0:N},\boldsymbol{X}^P_{1:N}}(\boldsymbol{x}_{0:N}, \boldsymbol{x}^P_{1:N}) &= f_{\boldsymbol{X}_0}(\boldsymbol{x}_0) \times \\ &\prod_{n=1}^{N} f_{\boldsymbol{X}_n|\boldsymbol{X}_{n-1},\boldsymbol{Y}_n}\big(\boldsymbol{x}_n \,|\, \boldsymbol{x}_{n-1}, \boldsymbol{y}^*_n\big)\; f_{\boldsymbol{X}_n|\boldsymbol{X}_{n-1}}\big(\boldsymbol{x}^P_n \,|\, \boldsymbol{x}_{n-1}\big). \end{aligned} \tag{4}$$

Using the convention that an empty density $f_{Y_\emptyset}$ evaluates to 1, we define

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

Denoting $\mathbb{E}_g$ for expectation for $(\boldsymbol{X}_{0:N}, \boldsymbol{X}^P_{1:N})$ having density $g_{\boldsymbol{X}_{0:N},\boldsymbol{X}^P_{1:N}}$, we have $f_{Y_{A_{u,n}}} = \mathbb{E}_g\big[\gamma_{A_{u,n}}\big]$ and thus

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

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,

$$\frac{\mathbb{E}_g\big[\gamma_{A^+_{u,n}}\big]}{\mathbb{E}_g\big[\gamma_{A_{u,n}}\big]} \approx \frac{\mathbb{E}_g\big[\gamma_{B^+_{u,n}}\big]}{\mathbb{E}_g\big[\gamma_{B_{u,n}}\big]}. \tag{5}$$

The conditional log likelihood estimate $\ell^{\mathrm{MC}}_{u,n}$ in ABF and ABF-IR come from replacing the expectations on the right hand side of (5) with averages over Monte Carlo replicates of simulations from the adapted process. To see that we expect the approximation in (5) to hold when dependence decays across spatiotemporal distance, we can write

$$\begin{aligned} \gamma_{A_{u,n}} &= \gamma_{B_{u,n}}\, \gamma_{B^c_{u,n}} \\ \gamma_{A^+_{u,n}} &= \gamma_{B^+_{u,n}}\, \gamma_{B^c_{u,n}}, \end{aligned}$$

where $B^c_{u,n}$ is the complement of $B_{u,n}$ in $A_{u,n}$. Under our assumptions, the term corresponding to $\gamma_{B^c_{u,n}}$ approximately cancels in the numerator and denominator of the right hand side of (5).

The localized likelihood estimate in ABF and ABF-IR has similar structure to UBF. However, ABF and ABF-IR additionally require 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 (as carried out by ABF) is vulnerable to COD. For a continuous time model, the use of intermediate resampling in ABF-IR is motivated by a result that this can reduce the COD, or avoid it entirely for an ideal guide function (Park and Ionides, 2020). Without intermediate resampling even an ideal proposal distribution does not avoid COD for a particle filter (Snyder et al., 2015). Assumptions B1–B4 below are analogous to A1–A4 and are non-asymptotic assumptions involving $\epsilon_{\mathrm{B1}} > 0$, $\epsilon_{\mathrm{B4}} > 0$ and $Q > 1$ which are required to hold uniformly over space and [time. Assumptions 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 polynomial in U and N when using an ideal guide function. This property is critical to enable ABF-IR to avoid COD. Since ABF has $S = 1$ it suffers from COD, albeit at an empirically slower rate than PF. The $\epsilon_{\mathrm{B6}}^{-3}$ 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) (details are provided in Sec. S15). 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 2 arises as the number of replicates increases.

Assumption B1. *There is an* $\epsilon_{\mathrm{B1}} > 0$*, 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* $|h(x)| \leq 1$*. Setting* $A = A_{u,n}$*,* $B = B_{u,n}$*,*

$f_A(x_A) = f_{Y_A|X_A}(y^*_A|x_A)$, *and* $f_B(x_B) = f_{Y_B|X_B}(y^*_B|x_B)$, *so that we have the identity*

$$f_{X_{u,n}|Y_A}(x|y^*_A) = \frac{\mathbb{E}_g\big[f_A\big(X^P_A\big) f_{X_{u,n}|X_{A^{[n]}},\boldsymbol{X}_{n-1}}\big(x|X^P_{A^{[n]}},\boldsymbol{X}_{n-1}\big)\big]}{\mathbb{E}_g\big[f_A\big(X^P_A\big)\big]},$$

we require that

$$\left|\int h(x)\left\{\frac{\mathbb{E}_g\big[f_A\big(X^P_A\big) f_{X_{u,n}|X_{A^{[n]}},\boldsymbol{X}_{n-1}}\big(x|X^P_{A^{[n]}},\boldsymbol{X}_{n-1}\big)\big]}{\mathbb{E}_g\big[f_A\big(X^P_A\big)\big]} - \frac{\mathbb{E}_g\big[f_B\big(X^P_B\big) f_{X_{u,n}|X_{B^{[n]}},\boldsymbol{X}_{n-1}}\big(x|X^P_{B^{[n]}},\boldsymbol{X}_{n-1}\big)\big]}{\mathbb{E}_g\big[f_B\big(X^P_B\big)\big]}\right\} dx\right| < \epsilon_{\mathrm{B1}}.$$

Assumption B2. *The bound* $\sup_{u\in 1:U, n\in 1:N}\left|B^+_{u,n}\right| \le b$ *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. *We use subscripts of* g *to denote marginal and conditional densities derived from (4). 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*

$$\left|g_{X^P_{B_{\tilde{u},\tilde{n}}\cup B_{u,n}}} - g_{X^P_{B_{\tilde{u},\tilde{n}}}}\, g_{X^P_{B_{u,n}}}\right| < (1/2)\,\epsilon_{\mathrm{B4}}\; g_{X^P_{B_{\tilde{u},\tilde{n}}\cup B_{u,n}}}$$

$$\left|g_{X^P_{B_{\tilde{u},\tilde{n}}}|\boldsymbol{X}_{0:N}}\, g_{X^P_{B_{u,n}}|\boldsymbol{X}_{0:N}} - g_{X^P_{B_{\tilde{u},\tilde{n}}\cup B_{u,n}}|\boldsymbol{X}_{0:N}}\right| < (1/2)\,\epsilon_{\mathrm{B4}}\; g_{X^P_{B_{\tilde{u},\tilde{n}}\cup B_{u,n}}|\boldsymbol{X}_{0:N}}$$

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

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

$$\left|g_{X_D|\boldsymbol{X}_{n-d-K}}(x_D\,|\,\boldsymbol{x}^{(1)}_{n-d-K}) - g_{X_D|\boldsymbol{X}_{n-d-K}}(x_D\,|\,\boldsymbol{x}^{(2)}_{n-d-K})\right| < \epsilon_{\mathrm{B5}}\; g_{X_D|\boldsymbol{X}_{n-d-K}}(x_D\,|\,\boldsymbol{x}^{(1)}_{n-d-K})$$

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

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

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

Assumption B7. *For $1 \leq n \leq N$, the Monte Carlo random variable $X^A_{n,i}$ is independent of $w^M_{u,n,i,j}$ conditional on $X^A_{n-1,i}$.*

Theorem 2. *Let $\ell^{\,MC}$ 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, $\mathcal{I} \to \infty$, 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 $|\epsilon| < Q^2\epsilon_{\mathrm{B1}} + 2Q^{2b}\big(\epsilon_{\mathrm{B5}} + (K + d_{\max})\epsilon_{\mathrm{B6}}\big)$ and $V < Q^{4b}U^2N^2$ such that*

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

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

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

Proof. A full proof is provided in Sec. S4. The extra work to prove Theorem 2 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. □

The theoretical results foreshadow our empirical observations (Sec. 4) 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.

4 Examples

We compare the performance of the three bagged filters (UBF, ABF and ABF-IR) against each other and against alternative plug-and-play approaches. The plug-and-play property facilitates numerical implementation for general classes of models, and all the algorithms and models under consideration are implemented in the R packages `pomp` (King et al., 2016) and `spatPomp` (Asfaw et al., 2021b). 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, 2009; 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 respect conservation laws and continuity or smoothness conditions obeyed by the dynamic model. We also compare with a guided intermediate resampling filter (GIRF, Park and Ionides, 2019), one of many variants of the particle filter designed to scale

to larger numbers of units than are possible with a basic particle filter.

First, in Sec. 4.1, we 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. Then, in Sec. 4.2, we develop a model for measles transmission within and between cities. The measles 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. In Sec. 4.3 we demonstrate an extension from likelihood evaluation to likelihood maximization for the measles model. Additionally, in Sec. S8, we compare performance on the Lorenz-96 model, a highly coupled system used to test inference methods for geophysical applications.

4.1 Correlated Brownian motion

Suppose $\boldsymbol{X}(t) = \Omega \boldsymbol{W}(t)$ where $\boldsymbol{W}(t) = W_{1:U}(t)$ 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,

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

Set $t_n = n$ for $n = 0, 1, \ldots, N$ with initial value $\boldsymbol{X}(0) = \boldsymbol{0}$ 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 $N = 50$. 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

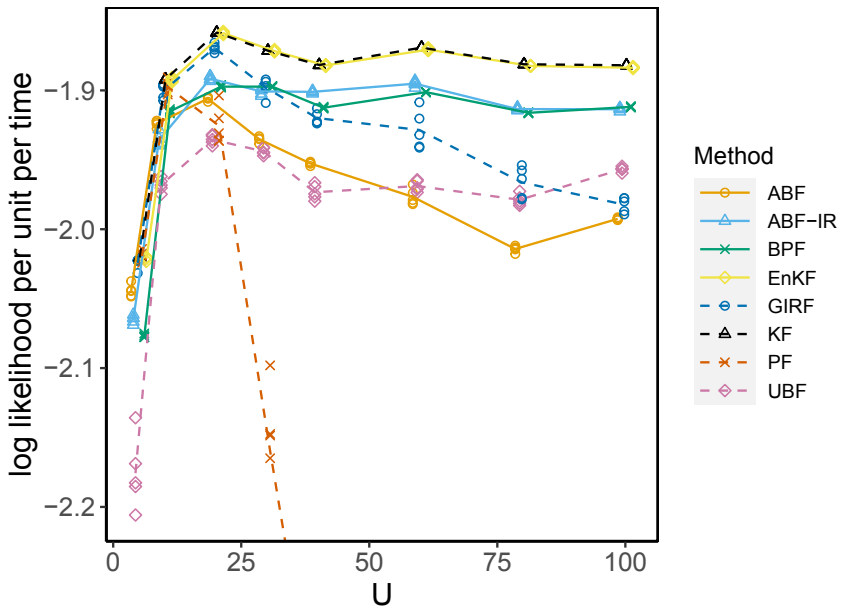

Figure 1: log likelihood estimates for a correlated Brownian motion model of various dimensions. UBF, ABF and ABF-IR are compared with a guided intermediate resampling filter (GIRF), standard particle filter (PF), block particle filter (BPF) and ensemble Kalman filter (EnKF). The exact likelihood was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

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., 2021a); for a Gaussian model, one can calculate an ideal guide function for GIRF but that was not used. PF works well for small values of U in Fig. 1 and 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.

4.2 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. We model the U largest cities in the UK, ordered in decreasing size so that $u = 1$ corresponds to London. We vary U to test methodologies on a hierarchy of filtering challenges. 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

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

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

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

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),

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

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 $\overline{d}$ 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 $\mu_{\bullet D} = 50$ 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 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

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

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 testing statistical tools rather than engaging directly in the scientific debate so we test methods on the simulated data.

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, 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

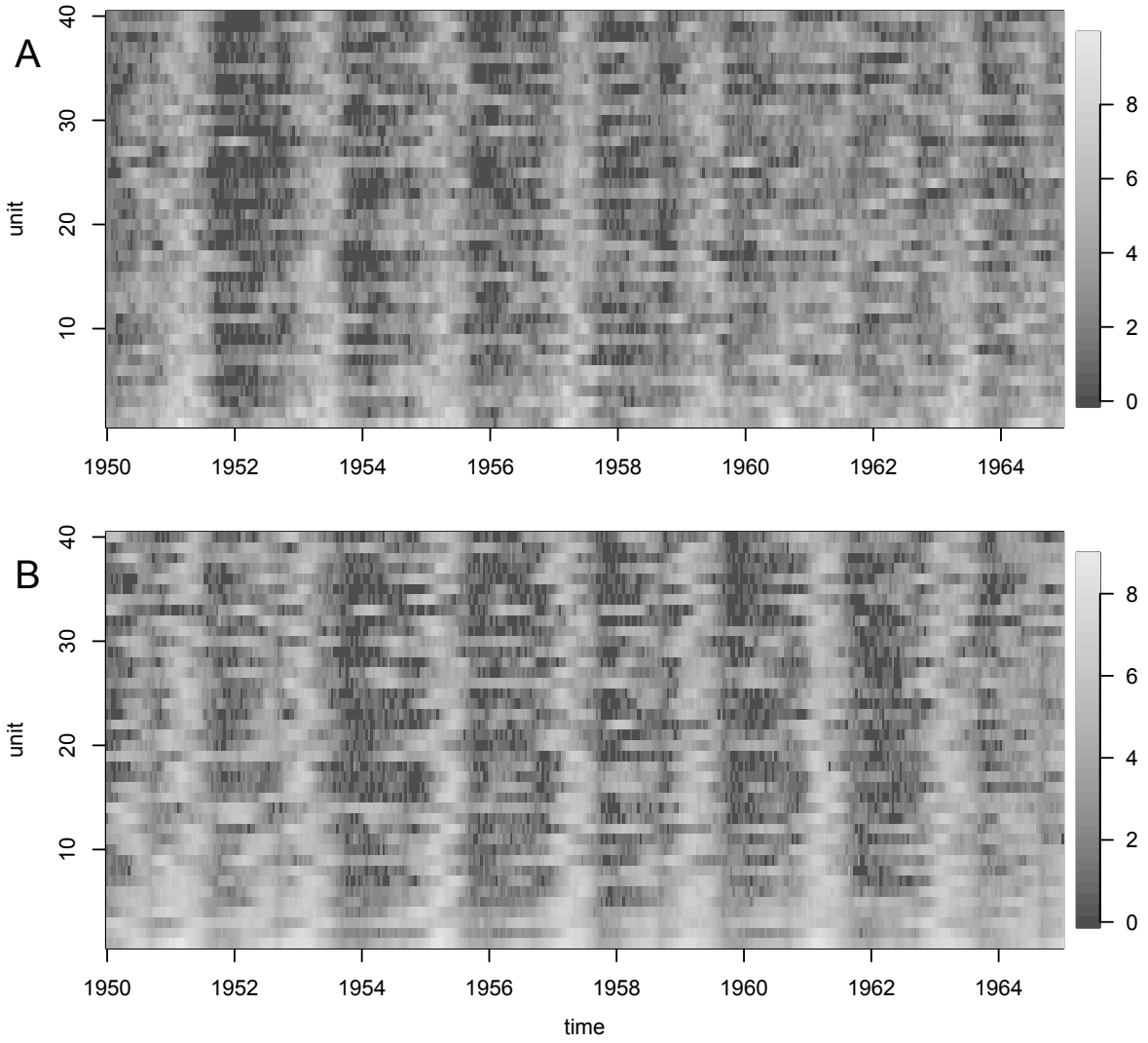

Figure 2: Log(reported cases + 1) for (A) the measles simulation used for the likelihood slice; (B) the corresponding UK measles data. The simulation shares the biennial pattern, with most but not all cities locked in phase most of the time.

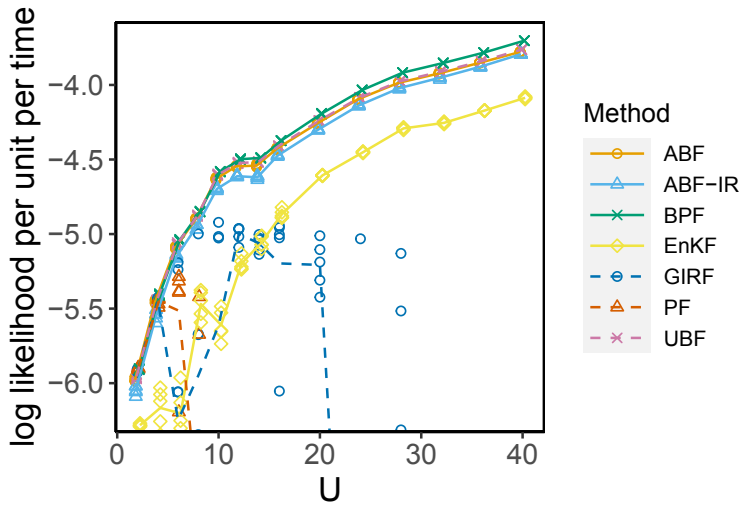

Figure 3: log likelihood estimates for simulated data from the measles model of various dimensions. UBF, ABF and ABF-IR are compared with a guided intermediate resampling filter (GIRF), a standard particle filter (PF), a block particle filter (BPF) and an ensemble Kalman filter (EnKF).

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. Thus, EnKF is relatively effective with small ensemble size but soon reaches the limit of its capabilities (Sec. S13). By contrast, the bagged filters and block particle filter perform substantially better than EnKF for larger ensemble size (Sec. S13). All the algorithms have

various other tuning parameters that could influence the results. Some investigations of alternatives are presented in Secs. S6, S7 and S12. 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. 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. Likelihood slices have less inferential value than likelihood profiles, but provide a computationally and conceptually simpler setting that can be insightful. Scientifically, the slices in Fig. 4 give 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.

4.3 Likelihood maximization and profile likelihood

Likelihood evaluation via filtering does not by itself enable parameter estimation for POMP models, however it provides a foundation for Bayesian and likelihood-based inference. In particular, filtering algorithms can be modified to carry out likelihood maximization by stochastically perturbing parameters in a sequence of filtering operations with decreasing perturbation variance (Ionides et al., 2015). We demonstrate this for the measles model in Fig. 5 using an iterated bagged filter algorithm which is fully described in Sec. S11.

Monte Carlo methods for computing and maximizing the log likelihood suffer from 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;

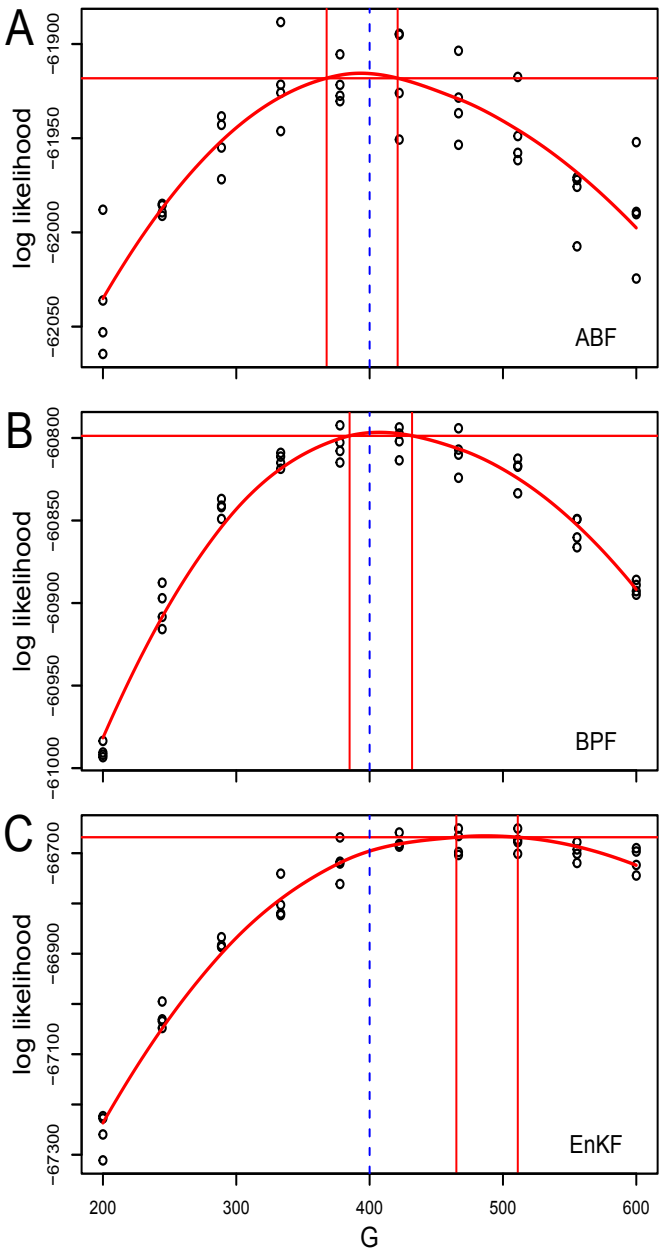

Figure 4: Likelihood slices varying the coupling parameter, for the measles model with $U = 40$ cities, computed via (A) ABF; (B) BPF; (C) EnKF. The solid perpendicular lines construct 95% Monte Carlo adjusted confidence intervals (Ionides et al., 2017). The true parameter value is identified by a blue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

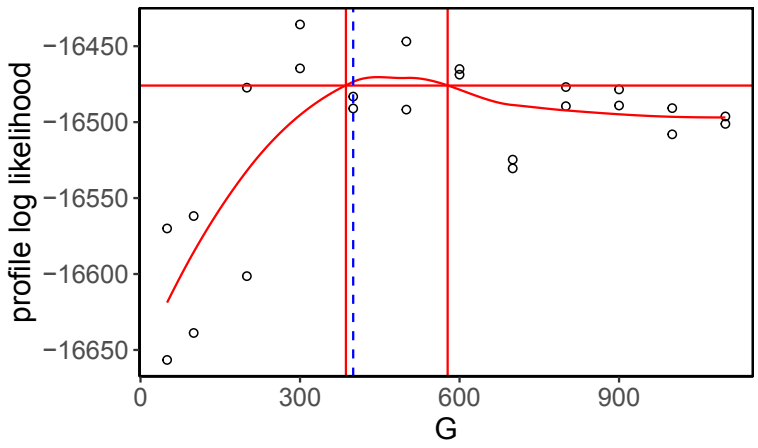

Figure 5: An iterated bagged filter used to maximize the likelihood, compute a profile likelihood, and hence construct a confidence interval. The profiling is carried out over the coupling parameter, G.

Ning et al., 2021). Fig. 5 constructs an MCAP 95% confidence interval for the coupling parameter, G, using an iterated unadapted bagged filter to maximize over the parameters, a, $\bar{\beta}$, σ_{SE}, ψ, μ_{EI} and μ_{IR}. This simulation study, carried out with $U = 20$ and $N = 208$, shows that G is identifiable via likelihood-based inference in the absence of assumptions about these parameters.

The likelihood estimate provided by bagged filters could be viewed as a composite likelihood (Varin et al., 2011) rather than an approximation to the likelihood. However, in situations where the likelihood approximation is found to be adequate, it is convenient to take advantage of the tools of likelihood-based inference.

5 Discussion

The nested loops used in the pseudocode for the bagged filters can be computed in various different orders to give mathematically equivalent results. There is scope for implementations to trade off memory, computation and communication by varying decisions on how the loops defined in the pseudocode are coded, including decisions on memory over-writing and

parallelization. This article focuses on the properties of the quantities calculated by the algorithms, leaving room for future research on implementation-specific considerations, though some supplementary discussion of memory-efficient implementation is given in Sec. S9.

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 population dynamics (Bretó, 2018; Pons-Salort and Grassly, 2018; de Cellès et al., 2018; Marino et al., 2019). This article develops a methodological extension motivated by the analysis of interacting biological populations. Similar challenges related to nonlinear non-Gaussian dynamic models arise in geophysical modeling. 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); Poterjoy et al. (2019) is perhaps closest to our approach, but the local weights of Poterjoy (2016); Poterjoy et al. (2019) are used to construct a localized Kalman gain which is motivated by a Gaussian approximation comparable to EnKF. EnKF arose originally via geophysical research (Evensen, 1994) and has since become used more widely for inference on SpatPOMP models (Katzfuss et al., 2020; Li et al., 2020). 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 they 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 true for the block particle filter, due to the indepdent resampling of the blocks (see Sec. S14). EnKF preserves linear constraints, since the filter procedure perturbs particles using a linear update rule, but 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., 2021b) 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 article is available as supplementary material.

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.

Asfaw, K., Ionides, E. L., and King, A. A. (2021a). `spatPomp`: R package for statistical inference for spatiotemporal partially observed Markov processes. `https://github.com/kidusasfaw/spatPomp`.

Asfaw, K., Park, J., Ho, A., King, A. A., and Ionides, E. L. (2021b). Statistical infer-

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

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.

Bentkus, V., Götze, F., and Tikhomoirov, A. (1997). Berry–Esseen bounds for statistics of weakly dependent samples. *Bernoulli*, 3(3):329–349.

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.

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.

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. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162.

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

Gordon, N., Salmond, D. J., and Smith, A. F. M. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE 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., 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.

Jirak, M. (2016). Berry–Esseen theorems under weak dependence. *The Annals of Probability*, 44(3):2024–2063.

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.

Ning, N., Ionides, E. L., and Ritov, Y. (2021). Scalable Monte Carlo inference and rescaled local asymptotic normality. *Bernoulli*, pre-published online.

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. (2019). Inference on high-dimensional implicit dynamic models using a guided intermediate resampling filter. *Arxiv:1708.08543v3*.

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.

Poterjoy, J., Wicker, L., and Buehner, M. (2019). Progress toward the application of a localized particle filter for numerical weather prediction. *Monthly Weather Review*, 147(4):1107–1126.

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.

Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, pages 5–42.

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.