Supplement to "Bagged filters for partially observed spatiotemporal systems"

E. L. Ionides, K. Asfaw, J. Park and A. A. King

September 25, 2020

Supplementary Content

S1 A generalization to models without latent unit structure

Variations of the algorithms in the main text apply when there is no latent unit structure. In this case, the observation vector $\boldsymbol{Y}_n = (Y_{1,n}, \ldots, Y_{U,n})$ consists of a collection of measurements on a general latent vector X_n. We may have the structure that $Y_{1,n}, \ldots, Y_{U,n}$ are conditionally independent given X_n, but even this is not essential to the approach. This is most readily seen in the context of the unadapted bagged filter, giving rise to the generalized unadapted bagged filter (G-UBF) algorithm defined as follows.

Algorithm G-UBF (Generalized unadapted bagged filter).

---

input:

Simulator for $f_{X_n|X_{n-1}}(x_n|x_{n-1})$

Evaluator for $f_{Y_{u,n}|X_n}(y^*_{u,n} \,|\, x_n)$

Number of bootstrap filters, $\mathcal{I}$

Neighborhood structure, $B_{u,n}$, for $u \in 1\!:\!U$ and $n \in 1\!:\!N$

Data, $y^*_{u,n}$ for $u \in 1\!:\!U$ and $n \in 1\!:\!N$

output:

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

For i in $1\!:\!\mathcal{I}$

  simulate $X^{\mathrm{sim}}_{n,i}$ from the dynamic model, for $n \in 1\!:\!N$

End For

Prediction weights, $w^P_{u,n,i} = f_{Y_{B_{u,n}}|X_{1:n}}(y^*_{B_{u,n}} \,|\, X^{\mathrm{sim}}_{1:n,i})$

Measurement weights, $w^M_{u,n,i} = f_{Y_{u,n}|X_n,Y_{B_{u,n}}}(y^*_{u,n} \,|\, X^{\mathrm{sim}}_{n,i}, y^*_{B_{u,n}})$

Conditional log likelihood estimate, $\ell^{\mathrm{MC}}_{u,n} = \log\left(\sum_{i=1}^{\mathcal{I}} w^M_{u,n,i}\, w^P_{u,n,i}\right) - \log\left(\sum_{\tilde{i}=1}^{\mathcal{I}} w^P_{u,n,\tilde{i}}\right)$

The algorithm G-UBF operates on an arbitrary POMP model. G-UBF therefore provides a potential approach to extending methodologies from SpatPOMP models to models that have some similarity to a SpatPOMP without formally meeting the definition. For example, there may be collections of interacting processes at different spatial scales in a spatiotemporal system. Alternatively, the potential outcomes of the latent process may vary between spatial units, such as when modeling interactions between terrestrial and aquatic ecosystems. We do not further explore G-UBF here.

S2 Adapted simulation for an Euler approximation

We investigate the adapted simulation process by considering a continuous-time limit where it becomes a diffusion process. We find that adapted simulation can effectively track the latent process when the measurement error is on an appropriate scale. However, when the measurement error is large compared to the latent process noise, adapted simulation can fail in situations where filtering succeeds. We work with a one-dimensional POMP model having a latent process constructed as an Euler approximation,

$$X_{n+1} \quad = \quad X_n + \mu(X_n)\delta + \sigma\sqrt{\delta}\epsilon_{n+1}, \tag{S1}$$

which provides a numerical solution to a one-dimensional stochastic differential equation,

$$dX(t) \quad = \quad \mu\big(X(t)\big)\,dt + \sigma\,dU(t),$$

where $\{U(t)\}$ is a standard Brownian motion. We will consider several different measurement processes.

S2.1 Measurement error on the same scale as the process noise

Here, we consider the measurement model

$$Y_{n+1} = \mu(X_n)\delta + \sigma\sqrt{\delta}\epsilon_{n+1} + \tau\sqrt{\delta}\,\eta_{n+1}. \tag{S2}$$

This is an approximation to the increment $Y(t+\delta)-Y(t)$ of a continuous time measurement model

$$dY(t) = dX(t) + \tau\,dV(t), \tag{S3}$$

where $\{V(t)\}$ is a standard Brownian motion independent of $\{U(t)\}$. The measurement model (S3) makes inference on $X(t)$ given $Y(t)$ a continuous time version of the filtering problem. A feature of this model is that $Y(t)$ does not directly track the level of the state, since the solution with initial conditions $Y(t_0) = X(t_0)$ and $V(t_0) = 0$ is

$$Y(t) = X(t) + \tau V(t).$$

The measurement error, $\tau V(t)$, has variance $\tau^2 t$ that increases with t. However, under appropriate conditions, information on changes in $\{X(t)\}$ obtained via $\{Y(t)\}$ are enough to track $X(t)$ indirectly via the filtering equations. For the POMP given by (S1) and (S2), we can calculate exactly the adapted simulation distribution $f_{X_{n+1}|Y_{n+1},X_n}$. It is convenient to work conditionally on X_n, allowing us to treat X_n and $\mu(X_n)$ as constants, with X_{n+1} and Y_{n+1} therefore being jointly normally distributed. A Gaussian distribution calculation then gives the conditional moments. First, we find

$$\begin{aligned}
\mathbb{E}\big[X_{n+1}|Y_{n+1},X_n\big] &= X_n + \mu(X_n)\delta + \mathbb{E}\big[\sigma\sqrt{\delta}\epsilon_{n+1}\,\big|\,\sigma\sqrt{\delta}\epsilon_{n+1} + \tau\sqrt{\delta}\eta_{n+1}\big] \\
&= X_n + \mu(X_n)\delta + \frac{\sigma^2}{\sigma^2+\tau^2}\big(\sigma\sqrt{\delta}\epsilon_{n+1} + \tau\sqrt{\delta}\eta_{n+1}\big) \\
&= X_n + \mu(X_n)\delta + \frac{\sigma^2}{\sigma^2+\tau^2}\big(Y_{n+1} - \mu(X_n)\delta\big).
\end{aligned}$$

Then,

$$\begin{aligned}\mathrm{Var}\big[X_{n+1}\,\big|\,Y_{n+1},X_n\big] &= \mathrm{Var}\big[\sigma\sqrt{\delta}\epsilon_{n+1}\,\big|\,\sigma\sqrt{\delta}\epsilon_{n+1}+\tau\sqrt{\delta}\eta_{n+1}\big]\\ &= \sigma^2\delta-\frac{\sigma^4\delta^2}{\sigma^2\delta+\tau^2\delta}\\ &= \delta\frac{\sigma^2\tau^2}{\sigma^2+\tau^2}.\end{aligned}$$

Call the adapted simulation process $\{A_n\}$, defined conditionally on $\{Y_n\}$. We see from the above calculation that

$$\begin{aligned}A_{n+1} &= A_n+\mu(A_n)\delta+\frac{\sigma^2}{\sigma^2+\tau^2}\Big(\mu(X_n)\delta+\sigma\sqrt{\delta}\,\epsilon_{n+1}+\tau\sqrt{\delta}\,\eta_{n+1}-\mu(A_n)\,\delta\Big)\\ &\qquad+\frac{\sigma\tau}{\sqrt{\sigma^2+\tau^2}}\sqrt{\delta}\,\zeta_{n+1}\end{aligned}$$

where $\{\zeta_n\}$ is an iid standard normal sequence independent of $\{\epsilon_n,\eta_n\}$. To study how well the adapted simulation tracks $\{X_n\}$, we subtract X_{n+1} from both sides to get

$$\begin{aligned}[A_{n+1}-X_{n+1}] &= [A_n-X_n]+[\mu(A_n)-\mu(X_n)]\delta-\sigma\sqrt{\delta}\,\epsilon_{n+1}\\ &\quad+\frac{\sigma^2}{\sigma^2+\tau^2}\Big([\mu(X_n)-\mu(A_n)]\delta+\sigma\sqrt{\delta}\,\epsilon_{n+1}+\tau\sqrt{\delta}\,\eta_{n+1}\Big)+\frac{\sigma\tau}{\sqrt{\sigma^2+\tau^2}}\sqrt{\delta}\,\zeta_{n+1}\\ &= [A_n-X_n]+\frac{2\sigma^2+\tau^2}{\sigma^2+\tau^2}[\mu(X_n)-\mu(A_n)]\delta\\ &\quad+\frac{\sigma^2\tau\sqrt{\delta}\eta_{n+1}-\sigma\tau^2\sqrt{\delta}\epsilon_{n+1}}{\sigma^2+\tau^2}+\frac{\sigma\tau}{\sqrt{\sigma^2+\tau^2}}\sqrt{\delta}\,\zeta_{n+1}.\end{aligned}$$

$\{A_n\}$ tracks $\{X_n\}$ when the process $\{A_n-X_n\}$ is stable. This happens when $\mu(x)-\mu(y)$ is negative when x is sufficiently larger than y. For example, a stable autoregressive process with $\mu(x)=-ax$ gives a stable adapted filter process.

S2.2 Independent measurement error on a scale that gives a finite limiting amount of information about $X(t)$ from measurements on a unit time interval

We now consider the measurement model

$$\begin{aligned}Y_{n+1} &= X_{n+1}+\frac{\tau}{\sqrt{\delta}}\,\eta_{n+1}\\ &= X_n+\mu(X_n)\,\delta+\sigma\sqrt{\delta}\,\epsilon_{n+1}+\frac{\tau}{\sqrt{\delta}}\eta_{n+1},\end{aligned}\tag{S4}$$

where $\{\epsilon_n,\eta_n\}$ is a collection of independent standard normal random variables. The conditional mean is now

$$\begin{aligned}\mathbb{E}\big[X_{n+1}|Y_{n+1},X_n\big]. &= X_n+\mu(X_n)\delta+\mathbb{E}\Big[\sigma\sqrt{\delta}\epsilon_{n+1}\,\Big|\,\sigma\sqrt{\delta}\epsilon_{n+1}+\frac{\tau}{\sqrt{\delta}}\eta_{n+1}\Big]\\ &= X_n+\mu(X_n)\delta+\frac{\sigma^2\delta}{\sigma^2\delta+\tau^2/\delta}\big(\sigma\sqrt{\delta}\epsilon_{n+1}+\tau\sqrt{\delta}\,\eta_{n+1}\big)\end{aligned}\tag{S5}$$

Using (S4) and (S5) gives

$$\mathbb{E}\big[X_{n+1}|Y_{n+1},X_n\big] \;=\; X_n+\mu(X_n)\delta+\frac{\sigma^2\delta^2}{\sigma^2\delta^2+\tau^2}\Big(Y_{n+1}-X_n-\mu(X_n)\,\delta\Big).$$

In the limit as $\delta\to 0$, the contribution from the measurement is order δ^2 and is therefore negligible. Although the observation process is meaningfully informative about the latent process, the adapted simulation fails to track the latent process in this limit. Intuitively, this is because the adapted simulation is trying to track differences in the latent process, but for this model the signal to noise ratio for the difference in each interval of length δ tends to zero.

S2.3 Independent measurements of the latent process with measurement error on a scale that gives a useful adapted process as $\delta\to 0$

We now consider the measurement model

$$\begin{aligned} Y_{n+1} &= X_{n+1}+\tau\eta_{n+1} \\ &= X_n+\mu(X_n)\delta+\sigma\sqrt{\delta}\epsilon_{n+1}+\tau\eta_{n+1}. \end{aligned} \tag{S6}$$

The conditional mean is now

$$\begin{aligned} \mathbb{E}\big[X_{n+1}|Y_{n+1},X_n\big]. &= X_n+\mu(X_n)\,\delta+\mathbb{E}\big[\sigma\sqrt{\delta}\epsilon_{n+1}\,\big|\,\sigma\sqrt{\delta}\epsilon_{n+1}+\tau\eta_{n+1}\big] \\ &= X_n+\mu(X_n)\delta+\frac{\sigma^2\delta}{\sigma^2\delta+\tau^2}\big(\sigma\sqrt{\delta}\epsilon_{n+1}+\tau\eta_{n+1}\big) \end{aligned} \tag{S7}$$

Using (S6) and (S7) gives

$$\begin{aligned} \mathbb{E}\big[X_{n+1}|Y_{n+1},X_n\big] &= X_n+\mu(X_n)\delta+\frac{\sigma^2\delta}{\sigma^2\delta+\tau^2}\Big(Y_{n+1}-X_n-\mu(X_n)\delta\Big) \\ &= X_n+\mu(X_n)\delta+\frac{\sigma^2}{\tau^2}\delta\Big(Y_{n+1}-X_n-\mu(X_n)\,\delta\Big)+o(\delta) \end{aligned}$$

In the limit as $\delta\to 0$, the adapted simulation has a diffusive drift toward the value of the latent process.

For disease models, incidence data can arguably be considered as noisy measurements of the change of a state variable (number of susceptibles) that is not directly measured. This could correspond to a situation where the measurement error is on the same scale as the process noise (Subsection S2.1). Alternatively, we could think of weekly aggregated incidence as a noisy measurement of the infected class, in which case the measurement error could match the scaling in Subsection S2.3.

The model in Subsection S2.2 is a cautionary tale, warning us against carrying out adapted simulation on short time intervals. An interpretation is that one should not carry out adapted simulation unless a reasonable amount of information has accrued. When each observation has low information, a particle filter may enable solution to the filtering problem without particle depletion. It is when the data are highly informative that the curse of dimensionality makes basic particle filters ineffective, opening up demand for alternative methods.

We are now in a better position to understand why it may be appropriate to keep many particle representations at intermediate timesteps while resampling down to a single representative at each

observation time, as ABF and ABF-IR do. We have seen that adaptive simulation can fail when observations occur frequently. Resampling down to a single particle too often can lose the ability for the adapted process to track the latent process. This implies that adapted simulation should not be relied upon more than necessary to ameliorate the curse of dimensionality: once proper importance sampling for filtering problem becomes tractable in a sufficiently small spatiotemporal neighborhood, one should maintain weighted particles on this spatiotemporal scale rather than resorting to adapted simulation.

S3 UBF convergence: Proof of Theorem 1

We consider a collection of models $f_{\boldsymbol{X}_{0:N},\boldsymbol{Y}_{1:N}}$ and data $\boldsymbol{y}^*_{1:N}$ defined for each (U, N). These models and datasets are not required to have any nesting relationship, so we do not insist that $X_{1,1}$ or $y^*_{1,1}$ should be the same for $(U, N) = (10, 10)$ as for $(U, N) = (100, 100)$. Formally, we define probability and expectation on a product space of the stochastic model and Monte Carlo outcomes. Monte Carlo quantities such as the output of the UBF, ABF and ABF-IR algorithms depend on the data but not on any random variables constructed under the model. As a consequence, we can use $\mathbb{E}$ to correspond both to expectations over Monte Carlo stochasticity (for Monte Carlo quantities) and model stochasticity (for random variables constructed under the model). We suppress discussion of measurability by assuming that all functions considered have appropriate measurability properties. We restate the assumptions and statement for Theorem 1.

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 $|h(x)| \le 1$, and any value of $x_{B^c_{u,n}}$,*

$$\left| \int h(x_{u,n}) f_{X_{u,n}|Y_{B_{u,n}},X_{B^c_{u,n}}}(x_{u,n} \mid y^*_{B_{u,n}}, x_{B^c_{u,n}})\, dx_{u,n} - \int h(x_{u,n}) f_{X_{u,n}|Y_{B_{u,n}}}(x_{u,n} \mid y^*_{B_{u,n}})\, dx_{u,n} \right| < \epsilon_{\mathrm{A1}}. \tag{S8}$$

We use the total variation bound in Assumption A1 via the following Proposition S1, which replaces conditioning on $X_{B^c_{u,n}}$ with conditioning on $Y_{B^c_{u,n}}$. The bound in (S9) could be used in place of Assumption A1.

Proposition S1. *Under the conditions of Assumption A1, (S8) implies*

$$\left| \int h(x_{u,n}) f_{X_{u,n}|Y_{A_{u,n}}}(x_{u,n} \mid y^*_{A_{u,n}})\, dx_{u,n} - \int h(x_{u,n}) f_{X_{u,n}|Y_{B_{u,n}}}(x_{u,n} \mid y^*_{B_{u,n}})\, dx_{u,n} \right| < \epsilon_{\mathrm{A1}} \tag{S9}$$

Proof. For notational compactness, we suppress the arguments $x_{u,n}$, $x_{B^c_{u,n}}$, $y^*_{A_{u,n}}$, $y^*_{B_{u,n}}$ matching the subscripts of conditional densities. Using the conditional independence of the measurements

given the latent process, we calculate

$$\begin{aligned}
&\left| \int h(x_{u,n}) f_{X_{u,n}|Y_{A_{u,n}}}\, dx_{u,n} - \int h(x_{u,n}) f_{X_{u,n}|Y_{B_{u,n}}}\, dx_{u,n} \right| \\
&\quad = \left| \int \left\{ \int h(x_{u,n}) f_{X_{u,n}|Y_{B_{u,n}},X_{B^c_{u,n}}}\, dx_{u,n} - \int h(x_{u,n}) f_{X_{u,n}|Y_{B_{u,n}}}\, dx_{u,n} \right\} f_{X_{B^c_{u,n}}\,|\,Y_{A_{u,n}}}\, dx_{B^c_{u,n}} \right| \\
&\quad \leq \int \left| \int h(x_{u,n}) f_{X_{u,n}|Y_{B_{u,n}},X_{B^c_{u,n}}}\, dx_{u,n} - \int h(x_{u,n}) f_{X_{u,n}|Y_{B_{u,n}}}\, dx_{u,n} \right| f_{X_{B^c_{u,n}}\,|\,Y_{A_{u,n}}}\, dx_{B^c_{u,n}} \\
&\quad < \int \epsilon_{\mathrm{A1}}\, f_{X_{B^c_{u,n}}\,|\,Y_{A_{u,n}}}\, dx_{B^c_{u,n}} \quad = \quad \epsilon_{\mathrm{A1}}.
\end{aligned}$$

□

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*

$$\left| f_{X_{B^+_{\tilde{u},\tilde{n}}}|X_{B^+_{u,n}}} - f_{X_{B^+_{\tilde{u},\tilde{n}}}} \right| < \epsilon_{\mathrm{A4}}\, f_{X_{B^+_{\tilde{u},\tilde{n}}}} \tag{S10}$$

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

Assumption A4 is needed only to ensure that the variance bound in Theorem 1 is essentially $O(UN)$ rather than $O(U^2N^2)$. Both these rates avoid the exponentially increasing variance characterizing the curse of dimensionality. Lower variance than $O(UN)$ cannot be anticipated for any sequential Monte Carlo method since the log likelihood estimate can be written as a sum of UN terms each of which involves its own sequential Monte Carlo calculation.

Theorem 1. *Let* $\ell^{\,MC}$ *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*

$$\mathcal{I}^{1/2} \big[\ell^{\,MC} - \ell - \epsilon UN \big] \xrightarrow[\mathcal{I}\to\infty]{d} \mathcal{N}\big[0, V\big], \tag{S11}$$

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*

$$V < Q^{4b}\, UN \big(c + \epsilon_{\mathrm{A4}}\, (UN - c) \big). \tag{S12}$$

Proof. Suppose the quantities $w^M_{u,n,i}$ and $w^P_{u,n,i}$ constructed in Algorithm UBF are considered i.i.d. replicates of jointly defined random variables $w^M_{u,n}$ and $w^P_{u,n}$, for each $(u,n) \in 1\!:\!U \times 1\!:\!N$. Also, write

$$\Delta^{MP}_{u,n} = \frac{1}{\sqrt{\mathcal{I}}} \sum_{i=1}^{\mathcal{I}} \big(w^M_{u,n,i} w^P_{u,n,i} - \mathbb{E}[w^M_{u,n} w^P_{u,n}]\big), \qquad \Delta^{P}_{u,n} = \frac{1}{\sqrt{\mathcal{I}}} \sum_{i=1}^{\mathcal{I}} \big(w^P_{u,n,i} - \mathbb{E}[w^P_{u,n}]\big),$$

Then, using the delta method (e.g., Section 2.5.3 in Liu (2001)) we find

$$\begin{aligned} \ell^{\mathrm{MC}}_{u,n} &= \log\left(\frac{\sum_{i=1}^{\mathcal{I}} w^M_{u,n,i} w^P_{u,n,i}}{\sum_{i=1}^{\mathcal{I}} w^P_{u,n,i}}\right) \\ &= \log\Big(\mathbb{E}[w^M_{u,n} w^P_{u,n}] + \mathcal{I}^{-1/2}\Delta^{MP}_{u,n}\Big) - \log\Big(\mathbb{E}[w^P_{u,n}] + \mathcal{I}^{-1/2}\Delta^{P}_{u,n}\Big) \\ &= \log\left(\frac{\mathbb{E}[w^M_{u,n} w^P_{u,n}]}{\mathbb{E}[w^P_{u,n}]}\right) + \mathcal{I}^{-1/2}\left(\frac{\Delta^{MP}_{u,n}}{\mathbb{E}[w^M_{u,n} w^P_{u,n}]} - \frac{\Delta^{P}_{u,n}}{\mathbb{E}[w^P_{u,n}]}\right) + o_P\big(\mathcal{I}^{-1/2}\big) \end{aligned} \tag{S13}$$

The joint distribution of $\big\{(\Delta^{MP}_{u,n}, \Delta^{P}_{u,n}), (u,n) \in 1\!:\!U \times 1\!:\!N\big\}$ follows a standard central limit theorem as $\mathcal{I} \to \infty$. Each term has mean zero, with covariances uniformly bounded over $(u,n,\tilde{u},\tilde{n})$ due to Assumption A3. Specifically,

$$\mathrm{Var}\begin{pmatrix} \Delta^{MP}_{u,n} \\ \Delta^{P}_{u,n} \\ \Delta^{MP}_{\tilde{u},\tilde{n}} \\ \Delta^{P}_{\tilde{u},\tilde{n}} \end{pmatrix} = \begin{pmatrix} \mathrm{Var}(w^M_{u,n} w^P_{u,n}) & \mathrm{Cov}(w^M_{u,n} w^P_{u,n}, w^P_{u,n}) & \mathrm{Cov}(w^M_{u,n} w^P_{u,n}, w^M_{\tilde{u},\tilde{n}} w^P_{\tilde{u},\tilde{n}}) & \mathrm{Cov}(w^M_{u,n} w^P_{u,n}, w^P_{\tilde{u},\tilde{n}}) \\ \mathrm{Cov}(w^M_{u,n} w^P_{u,n}, w^P_{u,n}) & \mathrm{Var}(w^P_{u,n}) & \mathrm{Cov}(w^P_{u,n}, w^M_{\tilde{u},\tilde{n}} w^P_{\tilde{u},\tilde{n}}) & \mathrm{Cov}(w^P_{u,n}, w^P_{\tilde{u},\tilde{n}}) \\ \mathrm{Cov}(w^M_{u,n} w^P_{u,n}, w^M_{\tilde{u},\tilde{n}} w^P_{\tilde{u},\tilde{n}}) & \mathrm{Cov}(w^P_{u,n}, w^M_{\tilde{u},\tilde{n}} w^P_{\tilde{u},\tilde{n}}) & \mathrm{Var}(w^M_{\tilde{u},\tilde{n}} w^P_{\tilde{u},\tilde{n}}) & \mathrm{Cov}(w^M_{\tilde{u},\tilde{n}} w^P_{\tilde{u},\tilde{n}}, w^P_{\tilde{u},\tilde{n}}) \\ \mathrm{Cov}(w^M_{u,n} w^P_{u,n}, w^P_{\tilde{u},\tilde{n}}) & \mathrm{Cov}(w^P_{u,n}, w^P_{\tilde{u},\tilde{n}}) & \mathrm{Cov}(w^M_{\tilde{u},\tilde{n}} w^P_{\tilde{u},\tilde{n}}, w^P_{\tilde{u},\tilde{n}}) & \mathrm{Var}(w^P_{\tilde{u},\tilde{n}}) \end{pmatrix}$$

Note that

$$\begin{aligned} \log\left[\frac{\mathbb{E}[w^M_{u,n} w^P_{u,n}]}{\mathbb{E}[w^P_{u,n}]}\right] &= \log\left[\frac{\int f_{Y_{u,n}|X_{u,n}}(y^*_{u,n} \,|\, x_{u,n,i})\, f_{Y_{B_{u,n}}|X_{B_{u,n}}}(y^*_{B_{u,n}} \,|\, x_{B_{u,n}})\, f_{X_{B^+_{u,n}}}(x_{B^+_{u,n}})\, dx_{B^+_{u,n}}}{\int f_{Y_{B_{u,n}}|X_{B_{u,n}}}(y^*_{B_{u,n}} \,|\, x_{B_{u,n}})\, f_{X_{B_{u,n}}}(x_{B_{u,n}})\, dx_{B_{u,n}}}\right] \\ &= \log\big[f_{Y_{u,n}|Y_{B_{u,n}}}(y^*_{u,n} \,|\, y^*_{B_{u,n}})\big], \end{aligned}$$

where $B^+_{u,n} = B_{u,n} \cup (u,n)$. Now, define

$$\Delta^{\ell}_{u,n} = \left(\frac{\Delta^{MP}_{u,n}}{\mathbb{E}[w^M_{u,n} w^P_{u,n}]} - \frac{\Delta^{P}_{u,n}}{\mathbb{E}[w^P_{u,n}]}\right)$$

Summing over all $(u,n) \in 1\!:\!U \times 1\!:\!N$, we get

$$\sqrt{\mathcal{I}}\left(\ell^{\,\mathrm{MC}} - \sum_{(u,n) \in 1\,:\,U \times 1\,:\,N} \log f_{Y_{u,n}|Y_{B_{u,n}}}(y^*_{u,n} \,|\, y^*_{B_{u,n}})\right) = \sum_{(u,n) \in 1\,:\,U \times 1\,:\,N} \Delta^{\ell}_{u,n} + o(1). \tag{S14}$$

Now,

$$\mathrm{Cov}\big(\Delta^{\ell}_{u,n}, \Delta^{\ell}_{\tilde{u},\tilde{n}}\big) = \mathrm{Cov}\left(\frac{w^M_{u,n} w^P_{u,n}}{\mathbb{E}[w^M_{u,n} w^P_{u,n}]} - \frac{w^P_{u,n}}{\mathbb{E}[w^P_{u,n}]}, \frac{w^M_{\tilde{u},\tilde{n}} w^P_{\tilde{u},\tilde{n}}}{\mathbb{E}[w^M_{\tilde{u},\tilde{n}} w^P_{\tilde{u},\tilde{n}}]} - \frac{w^P_{\tilde{u},\tilde{n}}}{\mathbb{E}[w^P_{\tilde{u},\tilde{n}}]}\right).$$

Since

$$\left| \frac{w^M_{u,n} w^P_{u,n}}{\mathbb{E}[w^M_{u,n} w^P_{u,n}]} - \frac{w^P_{u,n}}{\mathbb{E}[w^P_{u,n}]} \right| < Q^{2b},$$

we have

$$\left|\mathrm{Cov}\big(\Delta^\ell_{u,n}, \Delta^\ell_{\tilde{u},\tilde{n}}\big)\right| < Q^{4b},$$

implying that

$$\mathrm{Var}\left(\sum_{(u,n)\in 1:U\times 1:N} \Delta^\ell_{u,n} \right) < Q^{4b} U^2 N^2. \tag{S15}$$

If, in addition, (u,n) and $(\tilde{u},\tilde{n}) \in A_{u,n}$ are sufficiently separated in the sense of Assumption A4, then Lemma S1 shows that Assumption A4 implies

$$\left|\mathrm{Cov}\big(\Delta^\ell_{u,n}, \Delta^\ell_{\tilde{u},\tilde{n}}\big)\right| < \epsilon_{\mathrm{A4}} Q^{4b}.$$

The number of insufficiently separated neighbors to (u,n) is bounded by c, and so we obtain

$$\mathrm{Var}\left(\sum_{(u,n)\in S} \Delta^\ell_{u,n} \right) < Q^{4b}\, UN\big(c + \epsilon_{\mathrm{A4}}\,(UN - c)\big). \tag{S16}$$

Now we proceed to bound the bias in the Monte Carlo central limit estimator of ℓ. Putting $h(x_{u,n}) = f_{Y_{u,n}|X_{u,n}}(y^*_{u,n} \,|\, x_{u,n})$ into Assumption A1, using Assumption A3, gives

$$\left| f_{Y_{u,n}|Y_{B_{u,n}}}(y^*_{u,n} \,|\, y^*_{B_{u,n}}) - f_{Y_{u,n}|Y_{A_{u,n}}}(y^*_{u,n} \,|\, y^*_{A_{u,n}}) \right| < \epsilon_{\mathrm{A1}} Q.$$

Noting that

$$|a-b| < \delta, \quad a > Q^{-1} \text{ and } b > Q^{-1} \text{ implies } |\log(a) - \log(b)| < \delta Q, \tag{S17}$$

we find

$$\left| \log f_{Y_{u,n}|Y_{B_{u,n}}}(y^*_{u,n} \,|\, y^*_{B_{u,n}}) - \log f_{Y_{u,n}|Y_{A_{u,n}}}(y^*_{u,n} \,|\, y^*_{A_{u,n}}) \right| < \epsilon_{\mathrm{A1}} Q^2. \tag{S18}$$

Summing over (u,n), we get

$$\left| \ell - \sum_{(u,n)\in S} \log f_{Y_{u,n}|Y_{B_{u,n}}}(y^*_{u,n} \,|\, y^*_{B_{u,n}}) \right| < \epsilon_{\mathrm{A1}} Q^2 UN. \tag{S19}$$

Together, the results in [S14], [S15], [S16] and [S19] confirm the assertions of the theorem. □

Lemma S1. *Suppose U and V are random variables with joint density satisfying*

$$\left| f_{V|U}(v \,|\, u) - f_V(v) \right| < \epsilon f_V(v). \tag{S20}$$

Suppose $|g(U)| < a$ and $|h(V)| < b$ for some real-valued function g and h. Then, $\mathrm{Cov}\big(g(U), h(V)\big) < ab\epsilon$.

Proof. The result is obtained by direct calculation, as follows.

$$\begin{aligned}
\mathrm{Cov}\big(g(U),h(V)\big) &= \mathbb{E}\Big[g(U)\,\mathbb{E}\Big[h(V)-\mathbb{E}[h(V)]\,\Big|\,U\Big]\Big] \\
&= \int\Big\{\int g(u)h(v)\big(f_{V|U}(v\,|\,u)-f_V(v)\big)\,dv\Big\}\,f_U(u)\,du \\
&< \int\Big\{\int ab\epsilon f_V(v)\,dv\Big\}\,f_U(u)\,du \\
&= ab\epsilon.
\end{aligned}$$

□

S4 ABF and ABF-IR convergence: Proof of Theorem 2

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,

$$g_{\boldsymbol{X}_{0:N},\boldsymbol{X}^P_{1:N}}(\boldsymbol{x}_{0:N},\boldsymbol{x}^P_{1:N}) = f_{\boldsymbol{X}_0}(\boldsymbol{x}_0)\prod_{n=1}^{N} f_{\boldsymbol{X}_n|\boldsymbol{X}_{n-1},\boldsymbol{Y}_n}\big(\boldsymbol{x}_n\,|\,\boldsymbol{x}_{n-1},\boldsymbol{y}^*_n\big)\;f_{\boldsymbol{X}_n|\boldsymbol{X}_{n-1}}\big(\boldsymbol{x}^P_n\,|\,\boldsymbol{x}_{n-1}\big). \quad \text{(S21)}$$

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

$$\gamma_B=\prod_{m=1}^{N} f_{Y_{B^{[m]}}|\boldsymbol{X}_{m-1}}\big(y^*_{B^{[m]}}\,|\,\boldsymbol{X}_{m-1}\big), \quad \text{(S22)}$$

using the convention that an empty density $f_{Y_\emptyset}$ evaluates to 1. If we denoting $\mathbb{E}_g$ for expectation for $(\boldsymbol{X}_{0:N},\boldsymbol{X}^P_{1:N})$ having density $g_{\boldsymbol{X}_{0:N},\boldsymbol{X}^P_{1:N}}$, (S22) can be written as

$$\gamma_B=\mathbb{E}_g\left[f_{Y_B|X_B}\big(y^*_B\,|\,X^P_B\big)\,\middle|\,\boldsymbol{X}_{0:N}\right],$$

so we have

$$\mathbb{E}_g\big[\gamma_B\big]=\mathbb{E}_g\left[f_{Y_B|X_B}\big(y^*_B|X^P_B\big)\right].$$

Two useful identities are

$$\begin{aligned}
f_{X_{u,n}|Y_{A_{u,n}}}\big(x_{u,n}|y^*_{A_{u,n}}\big) &= \frac{\mathbb{E}_g\Big[f_{Y_{A_{u,n}}|X_{A_{u,n}}}\big(y^*_{A_{u,n}}\,|\,X^P_{A_{u,n}}\big)\,f_{X_{u,n}|X^{[n]}_{A_{u,n}},\boldsymbol{X}_{n-1}}\big(x_{u,n}|X^P_{A^{[n]}_{u,n}},\boldsymbol{X}_{n-1}\big)\Big]}{\mathbb{E}_g\Big[f_{Y_{A_{u,n}}|X_{A_{u,n}}}\big(y^*_{A_{u,n}}|X^P_{A_{u,n}}\big)\Big]}, \\
f_{Y_{u,n}|Y_{A_{u,n}}}\big(y^*_{u,n}|y^*_{A_{u,n}}\big) &= \frac{\mathbb{E}_g\big[\gamma_{A^+_{u,n}}\big]}{\mathbb{E}_g\big[\gamma_{A_{u,n}}\big]}.
\end{aligned}$$

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 $|h(x)| \leq 1$*: 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* $f_B(x_B) = f_{Y_B|X_B}(y_B^*|x_B)$*,*

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

Assumption B2. *The bound* $\sup_{u\in 1:U, n\in 1:N} \left|B_{u,n}^+\right| \leq b$ *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,* $Q^{-1} < f_{Y_{u,n}|X_{u,n}}(y_{u,n}^* \,|\, x_{u,n}) < Q$*.*

Proposition S2. *Setting* $h(x) = f_{Y_{u,n}|X_{u,n}}(y_{u,n}^*|x)$*, assumptions B1 and B3 imply*

$$\left| \frac{\mathbb{E}_g\big[\gamma_{A_{u,n}^+}\big]}{\mathbb{E}_g\big[\gamma_{A_{u,n}}\big]} - \frac{\mathbb{E}_g\big[\gamma_{B_{u,n}^+}\big]}{\mathbb{E}_g\big[\gamma_{B_{u,n}}\big]} \right| < Q\epsilon. \tag{S23}$$

Proof. Using the non-negativity of all terms to justify interchange of integral and expectation,

$$\begin{aligned} &\int h(x)\mathbb{E}_g\Big[f_{Y_{B_{u,n}}|X_{B_{u,n}}}(y_{B_{u,n}}^*|X_{B_{u,n}}^P) f_{X_{u,n}|X_{B_{u,n}^{[n]}},\boldsymbol{X}_{n-1}}(x|X_{B_{u,n}^{[n]}}^P,\boldsymbol{X}_{n-1})\Big]dx \\ &= \mathbb{E}_g\Big[\int f_{Y_{u,n}|X_{u,n}}(y_{u,n}^*|x) f_{X_{u,n}|X_{B_{u,n}^{[n]}},\boldsymbol{X}_{n-1}}(x|X_{B_{u,n}^{[n]}}^P,\boldsymbol{X}_{n-1})dx \cdot f_{Y_{B_{u,n}}|X_{B_{u,n}}}(y_{B_{u,n}}^*|X_{B_{u,n}}^P)\Big] \end{aligned} \tag{S24}$$

But by the construction of g,

$$\begin{aligned} f_{X_{u,n}|X_{B_{u,n}^{[n]}},\boldsymbol{X}_{n-1}}(x|X_{B_{u,n}^{[n]}}^P,\boldsymbol{X}_{n-1}) &= g_{X_{u,n}^P|X_{B_{u,n}^{[n]}}^P,\boldsymbol{X}_{n-1}}(x|X_{B_{u,n}^{[n]}}^P,\boldsymbol{X}_{n-1}) \\ &= g_{X_{u,n}^P|X_{B_{u,n}}^P,\boldsymbol{X}_{n-1}}(x|X_{B_{u,n}}^P,\boldsymbol{X}_{n-1}). \end{aligned}$$

Thus (S24) becomes

$$\begin{aligned} &\mathbb{E}_g\Big[\int f_{Y_{u,n}|X_{u,n}}(y_{u,n}^*|x) g_{X_{u,n}^P|X_{B_{u,n}^P},\boldsymbol{X}_{n-1}}(x|X_{B_{u,n}}^P,\boldsymbol{X}_{n-1})dx \cdot f_{Y_{B_{u,n}}|X_{B_{u,n}}}(y_{B_{u,n}}^*|X_{B_{u,n}}^P)\Big] \\ &= \mathbb{E}_g\left[\mathbb{E}_g\Big[f_{Y_{u,n}|X_{u,n}}(y_{u,n}^*|X_{u,n}^P)|X_{B_{u,n}}^P,\boldsymbol{X}_{n-1}\Big] \cdot f_{Y_{B_{u,n}}|X_{B_{u,n}}}(y_{B_{u,n}}^*|X_{B_{u,n}}^P)\right] \\ &= \mathbb{E}_g\Big[f_{Y_{B_{u,n}^+}|X_{B_{u,n}^+}}(y_{B_{u,n}^+}^*|X_{B_{u,n}^+}^P)\Big] \\ &= \mathbb{E}_g\gamma_{B_{u,n}^+}. \end{aligned}$$

Applying the same argument for the special case of $B_{u,n} = A_{u,n}$, we substitute into Assumption B1 to complete the proof with the fact that $h < Q$. □

Assumption B4. *Define $g_{X_A|X_B}$ via (S21). 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*

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

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

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

The mixing of the adapted process in Assumption B4 replaces the mixing of the unconditional process in Assumption A4. Though mixing of the adapted process may be hard to check, one may suspect that the adapted process typically mixes more rapidly than the unconditional process. Assumption B4 is needed only to ensure that the variance bound in Theorem **??** is essentially $O(UN)$ rather than $O(U^2N^2)$. Either of these rates avoids the exponentially increasing variance characterizing the curse of dimensionality. Lower variance than $O(UN)$ cannot be anticipated for any sequential Monte Carlo method since the log likelihood estimate can be written as a sum of UN terms each of which involves its own sequential Monte Carlo calculation. The following Proposition S3 gives an implication of Assumption B4

Proposition S3. *Assumption B4 implies that, if $(\tilde{u}, \tilde{n}) \notin C_{u,n}$,*

$$\mathrm{Cov}_g\big(\gamma_{B_{u,n}}, \gamma_{B_{\tilde{u},\tilde{n}}}\big) < \epsilon_{\mathrm{B4}} Q^{|B_{u,n}|+|B_{\tilde{u},\tilde{n}}|}. \tag{S25}$$

Proof. Write $\gamma = \gamma_{B_{u,n}}$ and $\tilde{\gamma} = \gamma_{B_{\tilde{u},\tilde{n}}}$. Also, write $B = B_{u,n}$, $\tilde{B} = B_{\tilde{u},\tilde{n}}$ and $f_B(x^P_B) = f_{Y_B|X_B}(y^*_B \,|\, x^P_B)$. Then,

$$\begin{aligned} \mathbb{E}[\gamma\tilde{\gamma}] &= \int \left[\int f_B(x^P_B) g_{X^P_B|\boldsymbol{X}_{0:N}}(x^P_B \,|\, \boldsymbol{x}_{0:N})\, dx^P_B \right] \\ &\qquad \times \left[\int f_{\tilde{B}}(x^P_{\tilde{B}}) g_{X^P_{\tilde{B}}|\boldsymbol{X}_{0:N}}(x^P_{\tilde{B}} \,|\, \boldsymbol{x}_{0:N})\, dx^P_{\tilde{B}} \right] g_{\boldsymbol{X}_{0:N}}(\boldsymbol{x}_{0:N})\, d\boldsymbol{x}_{0:N} \\ &= \int\int f_B(x^P_B)\, f_{\tilde{B}}(x^P_{\tilde{B}}) \left\{ \int g_{X^P_B|\boldsymbol{X}_{0:N}}(x^P_B \,|\, \boldsymbol{x}_{0:N}) \right. \\ &\qquad \left. \times\, g_{X^P_{\tilde{B}}|\boldsymbol{X}_{0:N}}(x^P_{\tilde{B}} \,|\, \boldsymbol{x}_{0:N})\, g_{\boldsymbol{X}_{0:N}}(\boldsymbol{x}_{0:N})\, d\boldsymbol{x}_{0:N} \right\} dx^P_B\, dx^P_{\tilde{B}} \end{aligned} \tag{S26}$$

Putting the approximate conditional independence requirement of Assumption B4 into (S26), we have

$$\begin{aligned} \left| \mathbb{E}[\gamma\tilde{\gamma}] - \int f_B(x^P_B)\, f_{\tilde{B}}(x^P_{\tilde{B}})\, g_{X^P_B X^P_{\tilde{B}}|\boldsymbol{X}_{0:N}}(x^P_B, x^P_{\tilde{B}} \,|\, \boldsymbol{x}_{0:N})\, g_{\boldsymbol{X}_{0:N}}(\boldsymbol{x}_{0:N})\, dx^P_B\, dx^P_{\tilde{B}}\, d\boldsymbol{x}_{0:N} \right| & \\ < (1/2)\, \epsilon_{\mathrm{B4}}\, Q^{|B|+|\tilde{B}|}. & \end{aligned}$$

This gives

$$\left|\mathbb{E}[\gamma\tilde{\gamma}] - \int f_B(x_B^P)\, f_{\tilde{B}}(x_{\tilde{B}}^P)\, g_{X_B^P X_{\tilde{B}}^P}(x_B^P, x_{\tilde{B}}^P)\, dx_B^P\, dx_{\tilde{B}}^P\right| < (1/2)\,\epsilon_{\mathrm{B4}}\, Q^{|B|+|\tilde{B}|}. \tag{S27}$$

Then, using the approximate unconditional independence requirement of Assumption B4 combined with the triangle inequality, (S27) implies

$$\left|\mathbb{E}[\gamma\tilde{\gamma}] - \int f_B(x_B^P)\, f_{\tilde{B}}(x_{\tilde{B}}^P)\, g_{X_B^P}(x_B^P)\, g_{X_{\tilde{B}}^P}(x_{\tilde{B}}^P)\, dx_B^P\, dx_{\tilde{B}}^P\right| < \epsilon_{\mathrm{B4}}\, Q^{|B|+|\tilde{B}|}. \tag{S28}$$

We can rewrite (S28) as

$$\left|\, \mathbb{E}[\gamma\tilde{\gamma}] - \mathbb{E}[\gamma]\,\mathbb{E}[\tilde{\gamma}]\, \right| < \epsilon_{\mathrm{B4}}\, Q^{|B|+|\tilde{B}|}, \tag{S29}$$

proving the proposition. □

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 $D \subset (1\!:\!U) \times (n\!:\!n-d)$,*

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

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

Assumption B5 is needed to ensure the stability of the Monte Carlo approximation to the adapted process. It ensures that any error due to finite Monte Carlo sample size has limited consequences at sufficiently remote time points. One could instead propose a bound that decreases exponentially with K, but that is not needed for the current purposes. The following Proposition S4 is useful for taking advantage of Assumption B5.

Proposition S4. *Suppose that f is a non-negative function and that for some $\epsilon > 0$,*

$$|f(x) - f(x')| < \epsilon f(x')$$

holds for all x, x'. Then for any two probability distributions where the expectations are denoted by $\mathbb{E}_1$ and $\mathbb{E}_2$ and for any random variable X, we have

$$|\mathbb{E}_1 f(X) - \mathbb{E}_2 f(X)| \le \epsilon\, \mathbb{E}_2 f(X).$$

Proof. Let the two probability laws be denoted by P_1 and P_2. We have

$$\begin{aligned}|\mathbb{E}_1 f(X) - \mathbb{E}_2 f(X)| &= \left|\int f(x) P_1(dx) - \int f(x') P_2(dx')\right| \\ &\le \left|\int\int f(x) P_1(dx) P_2(dx') - \int\int f(x') P_1(dx) P_2(dx')\right|. \\ &\le \int\int |f(x) - f(x')| P_1(dx) P_2(dx') \\ &\le \int \epsilon f(x') P_2(dx') = \epsilon\, \mathbb{E}_2 f(X).\end{aligned}$$

□

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 $J > \mathcal{C}_0(U,N,S)/\epsilon^3_{\mathrm{B6}}$,*

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

Assumption B6 controls the Monte Carlo error for a single time interval on a single bootstrap replicate. In the case $S = 1$, ABF-IR becomes ABF and this assumption is one of many alternatives for bounding error from importance sampling. The purpose behind the selection of Assumption B6 is to draw on the results of Park and Ionides (2020) for intermediate resampling, and our assumption is a restatement of their Theorem 2. When $S = 1$, the curse of dimensionality for importance sampling has the consequence that C_0 grows exponentially with U. However, Park and Ionides (2020) showed that setting $S = U$ can lead to situations where $C_0(U,N,S)$ in Assumption B6 grows polynomially with U. Here, we do not place requirements concerning the dependence of C_0 on U, N and S since our immediate concern is a limit where $\mathcal{I}$ and J increase. Nevertheless, the numerical results are consistent with the theoretical and empirical results obtained for intermediate resampling in the context of particle filtering by Park and Ionides (2020).

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 $X^A_{n-1,i}$.*

The Monte Carlo conditional independence required by Assumption B7 would hold for ABF-IR if the guide variance $V_{u,n,i}$ were calculated using an independent set of guide simulations to those used for evaluating the measurement weights $w^M_{u,n,i,j}$. For numerical efficiency, the ABF-IR algorithm implemented here constructs a shared pool of simulations for both purposes rather than splitting the pool up between them, in the expectation that the resulting minor violation of Assumption B7 has negligible impact.

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*

$$\mathcal{I}^{1/2}\big[\ell^{\,MC} - \ell - \epsilon UN\big] \xrightarrow[\mathcal{I}\to\infty]{d} \mathcal{N}\big[0, V\big]. \tag{S30}$$

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

$$V < Q^{4b}NU\big\{c + \big(\epsilon_{\mathrm{B4}} + 3\epsilon_{\mathrm{B5}} + 4(K + d_{\max})\,\epsilon_{\mathrm{B6}}\big)\big(NU - c\big)\big\} \tag{S31}$$

Proof. First, we set up some notation. For $B_{u,n}$ and $w^M_{u,n,i,j}$ constructed by ABF-IR, define

$$\gamma^{MC,i}_{B_{u,n}} = \prod_{m=1}^{n}\left[\frac{1}{J}\sum_{j=1}^{J}\prod_{(\tilde{u},m)\in B^{[m]}_{u,n}} w^M_{\tilde{u},m,i,j}\right] \quad \text{and} \quad \bar{\gamma}^{MC}_{B_{u,n}} = \frac{1}{\mathcal{I}}\sum_{i=1}^{\mathcal{I}}\gamma^{MC,i}_{B_{u,n}}. \tag{S32}$$

The Monte Carlo conditional likelihoods output by ABF-IR can be written as

$$\ell^{\mathrm{MC}}_{u,n} = \log \bar{\gamma}^{\mathrm{MC}}_{B^+_{u,n}} - \log \bar{\gamma}^{\mathrm{MC}}_{B_{u,n}}. \tag{S33}$$

We proceed with a similar argument to the proof of Theorem 1. Since $\gamma^{MC,i}_{B_{u,n}}$ are i.i.d. for $i \in 1{:}\mathcal{I}$, we can suppose they are replicates of a Monte Carlo random variable $\gamma^{MC}_{B_{u,n}}$. We define

$$\Delta^+_{u,n} = \frac{1}{\sqrt{\mathcal{I}}} \sum_{i=1}^{\mathcal{I}} \big(\gamma^{MC,i}_{B^+_{u,n}} - \mathbb{E}\big[\gamma^{MC}_{B^+_{u,n}}\big]\big), \qquad \Delta_{u,n} = \frac{1}{\sqrt{\mathcal{I}}} \sum_{i=1}^{\mathcal{I}} \big(\gamma^{MC,i}_{B_{u,n}} - \mathbb{E}\big[\gamma^{MC}_{B_{u,n}}\big]\big).$$

The same calculation as (S13) gives

$$\ell^{\mathrm{MC}}_{u,n} \;=\; \log\left(\frac{\mathbb{E}\big[\gamma^{MC}_{B^+_{u,n}}\big]}{\mathbb{E}\big[\gamma^{MC}_{B_{u,n}}\big]}\right) + \mathcal{I}^{-1/2}\left(\frac{\Delta^+_{u,n}}{\mathbb{E}\big[\gamma^{MC}_{B^+_{u,n}}\big]} - \frac{\Delta_{u,n}}{\mathbb{E}\big[\gamma^{MC}_{B_{u,n}}\big]}\right) + o_P\big(\mathcal{I}^{-1/2}\big) \tag{S34}$$

The joint distribution of $\big\{(\Delta^+_{u,n}, \Delta_{u,n}), (u,n) \in 1{:}U \times 1{:}N\big\}$ follows a standard central limit theorem as $\mathcal{I} \to \infty$. Each term has mean zero, with variances and covariances uniformly bounded over $(u, n, \tilde{u}, \tilde{n})$ due to Assumption B3. From Proposition S2, using the same reasoning as (S18),

$$\left|\log\left(\frac{\mathbb{E}_g\big[\gamma_{A^+_{u,n}}\big]}{\mathbb{E}_g\big[\gamma_{A_{u,n}}\big]}\right) - \log\left(\frac{\mathbb{E}_g\big[\gamma_{B^+_{u,n}}\big]}{\mathbb{E}_g\big[\gamma_{B_{u,n}}\big]}\right)\right| < \epsilon_{\mathrm{B1}} Q^2. \tag{S35}$$

Now we use Lemma S2 and (S17) to obtain

$$\begin{aligned}
&\left|\log\left(\frac{\mathbb{E}_g\big[\gamma_{B^+_{u,n}}\big]}{\mathbb{E}_g\big[\gamma_{B_{u,n}}\big]}\right) - \log\left(\frac{\mathbb{E}\big[\gamma^{MC}_{B^+_{u,n}}\big]}{\mathbb{E}\big[\gamma_{B_{u,n}}\big]}\right)\right| \\
&\leq \left|\log \mathbb{E}_g\big[\gamma_{B^+_{u,n}}\big] - \log \mathbb{E}\big[\gamma_{B^+_{u,n}}\big]\right| + \left|\log \mathbb{E}_g\big[\gamma_{B_{u,n}}\big] - \log \mathbb{E}\big[\gamma_{B_{u,n}}\big]\right| \\
&< 2Q^{2b}\big(\epsilon_{\mathrm{B5}} + (K + d_{\max})\epsilon_{\mathrm{B6}}\big).
\end{aligned} \tag{S36}$$

The proof of the central limit result in (**??**) is completed by combining (S34), (S35) and (S36). To show (**??**) we check that $\Delta_{u,n}$ and $\Delta_{\tilde{u},\tilde{n}}$ are weakly correlated when (u,n) and $(\tilde{u},\tilde{n})$ are sufficiently separated. By the same reasoning as the proof of Theorem 1, it is sufficient to show that $\gamma^{MC}_{B_{u,n}}$ and $\gamma^{MC}_{B_{\tilde{u},\tilde{n}}}$ are weakly correlated. These Monte Carlo quantities approximate $\gamma_{B_{u,n}}(\boldsymbol{X}_{0:n-1})$ and $\gamma_{B_{\tilde{u},\tilde{n}}}(\boldsymbol{X}_{0:\tilde{n}-1})$ with $\boldsymbol{X}$ drawn from g. Let us suppose $n \geq \tilde{n}$, and write $d_{u,n} = n - \inf_{(v,m)\in B_{u,n}} m$. First, we consider the situation $n - \tilde{n} > K + d_{u,n}$, in which case we can use the Markov property to give

$$\mathrm{Cov}\big(\gamma^{MC}_{B_{u,n}}, \gamma^{MC}_{B_{\tilde{u},\tilde{n}}}\big) < \mathbb{E}\big[\gamma^{MC}_{B_{\tilde{u},\tilde{n}}}\big] \sup_{\boldsymbol{x}} \Big\{\mathbb{E}\big[\gamma^{MC}_{B_{u,n}} \big| \boldsymbol{X}^A_{n-d_{u,n}-K,1} = \boldsymbol{x}\big] - \mathbb{E}\big[\gamma^{MC}_{B_{u,n}}\big]\Big\} \tag{S37}$$

Then, the triangle inequality followed by applications of Assumption B5 and Lemma S2 gives

$$\begin{aligned}
&\Big|\mathbb{E}\big[\gamma^{MC}_{B_{u,n}}\big|\boldsymbol{X}^{A}_{n-d_{u,n}-K,1}=\boldsymbol{x}\big]-\mathbb{E}\big[\gamma^{MC}_{B_{u,n}}\big]\Big| \\
&\qquad\le \Big|\mathbb{E}_g\big[\gamma_{B_{u,n}}\big|\boldsymbol{X}_{n-d_{u,n}-K}=\boldsymbol{x}\big]-\mathbb{E}_g\big[\gamma_{B_{u,n}}\big]\Big| \\
&\qquad\qquad +\Big|\mathbb{E}\big[\gamma^{MC}_{B_{u,n}}\big|\boldsymbol{X}^{A}_{n-d_{u,n}-K,1}=\boldsymbol{x}\big]-\mathbb{E}_g\big[\gamma_{B_{u,n}}\big|\boldsymbol{X}_{n-d_{u,n}-K}=\boldsymbol{x}\big]\Big| \\
&\qquad\qquad +\Big|\mathbb{E}\big[\gamma^{MC}_{B_{u,n}}\big]-\mathbb{E}_g\big[\gamma_{B_{u,n}}\big]\Big| \\
&\qquad\le Q^b\Big(2\epsilon_{\mathrm{B5}}+2\big(K+d_{u,n}\big)\epsilon_{\mathrm{B6}}\Big) \qquad \text{(S38)}
\end{aligned}$$

Putting (S38) into (S37), we get

$$\mathrm{Cov}\big(\gamma^{MC}_{B_{u,n}},\gamma^{MC}_{B_{\tilde{u},\tilde{n}}}\big)<Q^{2b}\Big(2\epsilon_{\mathrm{B5}}+2\big(K+d_{u,n}\big)\epsilon_{\mathrm{B6}}\Big). \qquad \text{(S39)}$$

Now we address the situation $n-\tilde{n}\le K+d_{u,n}$. We apply Lemma S2 on the union $B_{u,n}\cup B_{\tilde{u},\tilde{n}}$ for which the temporal depth is bounded by $d\le K+d_{u,n}+d_{\tilde{u},\tilde{n}}$. This gives

$$\Big|\,\mathbb{E}\big[\gamma^{MC}_{B_{u,n}}\gamma^{MC}_{B_{\tilde{u},\tilde{n}}}\big]-\mathbb{E}_g\big[\gamma_{B_{u,n}}\gamma_{B_{\tilde{u},\tilde{n}}}\big]\,\Big|<Q^{2b}\Big((2K+d_{u,n}+d_{\tilde{u},\tilde{n}})\epsilon_{\mathrm{B6}}+\epsilon_{\mathrm{B5}}\Big). \qquad \text{(S40)}$$

From Proposition S3, if $(\tilde{u},\tilde{n})\notin C_{u,n}$,

$$\mathrm{Cov}_g\big(\gamma_{B_{u,n}},\gamma_{B_{\tilde{u},\tilde{n}}}\big)<\epsilon_{\mathrm{B4}}Q^{2b}. \qquad \text{(S41)}$$

Now, we establish that $\mathrm{Cov}\big(\gamma^{MC}_{B_{u,n}},\gamma^{MC}_{B_{\tilde{u},\tilde{n}}}\big)$ is close to $\mathrm{Cov}_g\big(\gamma_{B_{u,n}},\gamma_{B_{\tilde{u},\tilde{n}}}\big)$.

$$\begin{aligned}
&\Big|\mathrm{Cov}\big(\gamma^{MC}_{B_{u,n}},\gamma^{MC}_{B_{\tilde{u},\tilde{n}}}\big)-\mathrm{Cov}_g\big(\gamma_{B_{u,n}},\gamma_{B_{\tilde{u},\tilde{n}}}\big)\Big| \\
&\qquad\le \Big|\,\mathbb{E}\big[\gamma^{MC}_{B_{u,n}}\gamma^{MC}_{B_{\tilde{u},\tilde{n}}}\big]-\mathbb{E}_g\big[\gamma_{B_{u,n}}\gamma_{B_{\tilde{u},\tilde{n}}}\big]\,\Big| \\
&\qquad\qquad +\Big|\mathbb{E}\big[\gamma^{MC}_{B_{u,n}}\big]\big(\mathbb{E}\big[\gamma^{MC}_{B_{\tilde{u},\tilde{n}}}\big]-\mathbb{E}_g\big[\gamma_{B_{\tilde{u},\tilde{n}}}\big]\big)\Big| \\
&\qquad\qquad +\Big|\mathbb{E}\big[\gamma_{B_{\tilde{u},\tilde{n}}}\big]\big(\mathbb{E}\big[\gamma^{MC}_{B_{u,n}}\big]-\mathbb{E}_g\big[\gamma_{B_{u,n}}\big]\big)\Big| \\
&\qquad< Q^{2b}\Big((2K+d_{u,n}+d_{\tilde{u},\tilde{n}})\epsilon_{\mathrm{B6}}+\epsilon_{\mathrm{B5}}+2(\epsilon_{\mathrm{B5}}+(K+d_{\max})\epsilon_{\mathrm{B6}})\Big). \\
&\qquad< Q^{2b}\Big(3\epsilon_{\mathrm{B5}}+4(K+d_{\max})\,\epsilon_{\mathrm{B6}}\Big) \qquad \text{(S42)}
\end{aligned}$$

Using (S42) together with (S41) to bound the $UN(UN-c)$ off-diagonal covariance terms completes the derivation of (**??**).

□

Lemma S2. *Suppose Assumptions B3, B5, B6 and B7. Suppose the number of particles J exceeds the requirement for B6. If we write $d_B=\max_{(u_1,n_1),(u_2,n_2)\in B}|n_1-n_2|$ for $B\subset 1\!:\!U\times 1\!:\!N$, then for any B,*

$$\Big|\,\mathbb{E}\big[\gamma^{MC}_{B}\big|\boldsymbol{X}^{A}_{n-d_B-K,1}=\boldsymbol{x}\big]-\mathbb{E}_g\big[\gamma_B\big|\boldsymbol{X}_{n-d_B-K}=\boldsymbol{x}\big]\,\Big|<Q^{|B|}(K+d_B)\epsilon_{\mathrm{B6}},\quad\forall\boldsymbol{x}\in\mathbb{X}^U,$$

and

$$\Big|\,\mathbb{E}\big[\gamma^{MC}_{B}\big]-\mathbb{E}_g\big[\gamma_B\big]\,\Big|<Q^{|B|}(\epsilon_{\mathrm{B5}}+(K+d_B)\epsilon_{\mathrm{B6}}). \qquad \text{(S43)}$$

Proof. Suppose that $\max_{(u',n')\in B} n' = n$. Define $\eta_n(\boldsymbol{x}_n) = 1$ and, for $0 \le m \le n-1$,

$$\eta_m(\boldsymbol{x}_m) = \mathbb{E}_g\left[\prod_{k=m+1}^{n} \gamma_{B^{[k]}} \,\middle|\, \boldsymbol{X}_m = \boldsymbol{x}_m\right]. \tag{S44}$$

We have a recursive identity

$$\eta_m(\boldsymbol{X}_m) = \mathbb{E}_g\left[\gamma_{B^{[m+1]}}\, \eta_{m+1}(\boldsymbol{X}_{m+1}) \middle| \boldsymbol{X}_m\right]. \tag{S45}$$

By taking the expectation of (S44), we have

$$\mathbb{E}_g\big[\eta_0(\boldsymbol{X}_0)\big] = \mathbb{E}_g\big[\gamma_B\big]. \tag{S46}$$

Note that g has marginal density $f_{\boldsymbol{X}_0}$ for $\boldsymbol{X}_0$. We analyze an ABF-IR approximation to (S45). The function $\eta_{m+1}(\boldsymbol{x})$ is not in practice computationally available for evaluation via ABF-IR, but the recursion nevertheless leads to a useful bound. Let $\boldsymbol{X}^{\mathrm{A}}_{m+1}[j](\boldsymbol{x}_m)$ correspond to the variable $\boldsymbol{X}^{IR}_{m+1,S,1,j}$ constructed by ABF-IR conditional on $\boldsymbol{X}^{\mathrm{A}}_{m,1} = \boldsymbol{x}_m$. Equivalently, $\boldsymbol{X}^{\mathrm{A}}_{m+1}[j](\boldsymbol{x}_m)$ matches the variable $\boldsymbol{X}^{\mathrm{A}}_{m+1,1}$ in ABF-IR if the assignment $\boldsymbol{X}^{\mathrm{A}}_{m+1,1} = \boldsymbol{X}^{IR}_{m+1,S,1,1}$ is replaced by $\boldsymbol{X}^{\mathrm{A}}_{m+1,1} = \boldsymbol{X}^{IR}_{m+1,S,1,j}$ conditional on $\boldsymbol{X}^{\mathrm{A}}_{m,1} = \boldsymbol{x}_m$. We define an approximation error $e_m(\boldsymbol{x}_m)$ by

$$\eta_m(\boldsymbol{x}_m) = \frac{1}{J}\sum_{j=1}^{J} f_{Y_{B^{[m+1]}}|\boldsymbol{X}_m}(y^*_{B^{[m+1]}} \mid \boldsymbol{x}_m)\, \eta_{m+1}\big(\boldsymbol{X}^{\mathrm{A}}_{m+1}[j](\boldsymbol{x}_m)\big) + e_m(\boldsymbol{x}_m). \tag{S47}$$

From Assumptions B3 and B6, $\mathbb{E}\big|e_m(\boldsymbol{x}_m)\big| < \epsilon_{\mathrm{B6}}\, Q^{\left|B^{[m+1:n]}\right|}$ uniformly over $\boldsymbol{x}_m$, Thus, setting $r_m = \mathbb{E}|e_m(\boldsymbol{X}^A_{m,1})|$, we have

$$r_m < \epsilon_{\mathrm{B6}}\, Q^{\left|B^{[m+1:n]}\right|}. \tag{S48}$$

Now, setting $K' = K + d_{u,n}$, we commence to prove inductively that, for $n - K' \le m \le n$,

$$\left|\eta_{n-K'}(\boldsymbol{x}) - \mathbb{E}\Big[\eta_m(\boldsymbol{X}^A_{m,1}) \prod_{k=n-K'+1}^{m} f_{Y_{B[k]}|\boldsymbol{X}_{k-1}}(y^*_{B[k]} \mid \boldsymbol{X}^A_{k-1,1}) \Big| \boldsymbol{X}^A_{n-K',1} = \boldsymbol{x}\Big]\right| < (m-n+K')\epsilon_{\mathrm{B6}}\, Q^{|B|}. \tag{S49}$$

First, suppose that (S49) holds for m. From (S47) and (S48),

$$\left|\eta_m(\boldsymbol{x}_m) - \mathbb{E}\left[\frac{1}{J}\sum_{j=1}^{J} f_{Y_{B^{[m+1]}}|\boldsymbol{X}_m}(y^*_{B^{[m+1]}} \mid \boldsymbol{x}_m)\, \eta_{m+1}\big(\boldsymbol{X}^{\mathrm{A}}_{m+1}[j](\boldsymbol{x}_m)\big)\right]\right| < \epsilon_{\mathrm{B6}}\, Q^{\left|B^{[m+1:n]}\right|}. \tag{S50}$$

Since the particles are exchangeable, the expectation of the mean of J particles can be replaced with the expectation of the first particle. Plugging in $\boldsymbol{x}_m = \boldsymbol{X}^A_{m,1}$ gives us

$$\Big|\eta_m(\boldsymbol{X}^A_{m,1}) - f_{Y_{B^{[m+1]}}|\boldsymbol{X}_m}\big(y^*_{B^{[m+1]}} \mid \boldsymbol{X}^A_{m,1}\big)\mathbb{E}\left[\eta_{m+1}\big(\boldsymbol{X}^A_{m+1,1}\big)\big|\boldsymbol{X}^A_{m,1}\right]\Big| < \epsilon_{\mathrm{B6}}\, Q^{\left|B^{[m+1:n]}\right|} \tag{S51}$$

Putting (S51) into (S49), for $m \le n$, and taking an iterated expectation with respect to $\boldsymbol{X}^A_{m,1}$, we find that (S49) holds also for $m+1$. Since (S49) holds trivially for $m = n - K'$, it holds for $n - K' \le m \le n$ by induction. Then, noting $\eta_n(\boldsymbol{x}) = 1$, we have from (S49) that

$$\left|\eta_{n-K'}(x) - \mathbb{E}\left[\prod_{k=n-K'+1}^{n} f_{Y_{B[k]}|\boldsymbol{X}_{k-1}}\big(y^*_{B^{[k]}} \mid \boldsymbol{X}^A_{k-1,1}\big) \middle| \boldsymbol{X}^A_{n-K',1} = x\right]\right| < K'\epsilon_{\mathrm{B6}}\, Q^{|B|}.$$

Integrating the above inequality over $\boldsymbol{x}$ with respect to the law of $\boldsymbol{X}^A_{n-K',1}$, we obtain

$$\left|\mathbb{E}[\eta_{n-K'}(\boldsymbol{X}^A_{n-K',1})] - \mathbb{E}\left[\prod_{k=n-K'+1}^{n} f_{Y_{B^{[k]}}|\boldsymbol{X}_{k-1}}\big(y^*_{B^{[k]}} \mid \boldsymbol{X}^A_{k-1,1}\big)\right]\right| < K'\epsilon_{\mathrm{B6}}\, Q^{|B|}. \tag{S52}$$

But under Assumption B7, we have

$$\mathbb{E}\left[\prod_{k=n-K'+1}^{n} f_{Y_{B^{[k]}}|\boldsymbol{X}_{k-1}}\big(y^*_{B^{[k]}} \mid \boldsymbol{X}^A_{k-1,1}\big)\right] = \mathbb{E}\big[\gamma^{MC}_B\big]. \tag{S53}$$

Assumption B5 says

$$|\eta_{n-K'}(\boldsymbol{x}^{(1)}_{n-K'}) - \eta_{n-K'}(\boldsymbol{x}^{(2)}_{n-K'})| < \epsilon_{\mathrm{B5}}\, \eta_{n-K'}(\boldsymbol{x}^{(2)}_{n-K'}). \tag{S54}$$

Application of Proposition S4 to (S54) gives

$$\Big|\mathbb{E}_g\big[\eta_{n-K'}(\boldsymbol{X}_{n-K'})\big] - \mathbb{E}\big[\eta_{n-K'}(\boldsymbol{X}^A_{n-K',1})\big]\Big| < \epsilon_{\mathrm{B5}}\, \mathbb{E}_g\big[\eta_{n-K'}(\boldsymbol{X}_{n-K'})\big] < \epsilon_{\mathrm{B5}}\, Q^{|B|}. \tag{S55}$$

Combining (S52), (S53), and (S55) completes the proof of Lemma S2. □

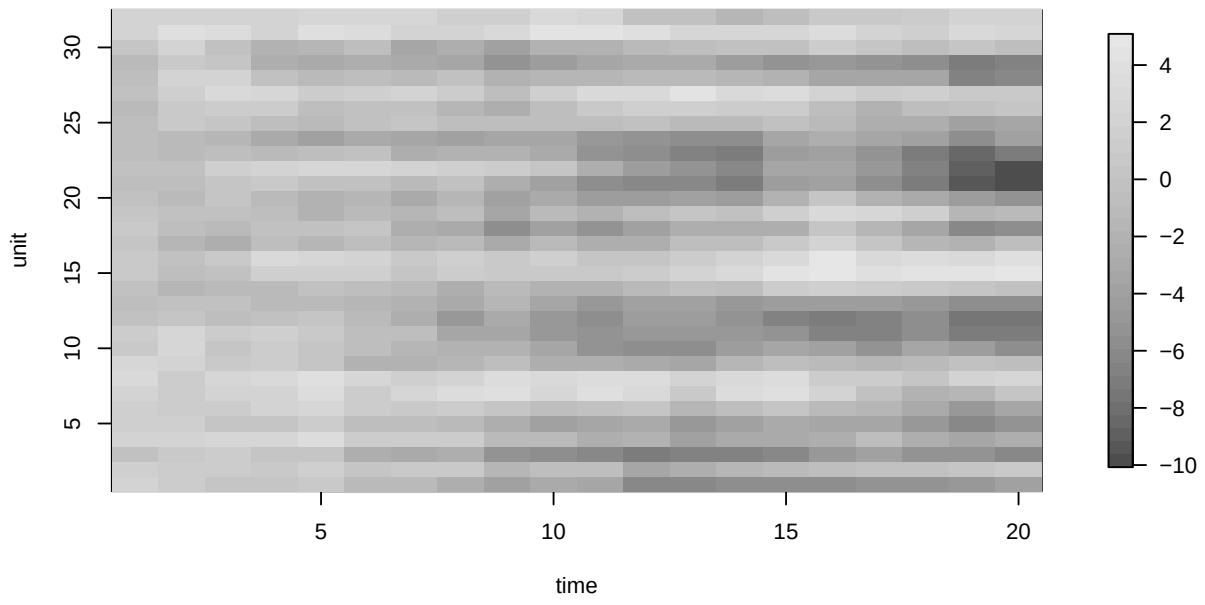

Figure S-1: correlated Brownian motion simulation used in the main text

S5 The correlated Brownian motion example

To help visualize the correlated Brownian motion model, Fig. S-1 shows one of the simulations used for the results in Figure 3 of the main text. Table S-1 gives the algorithmic settings used for the filters and corresponding computational resource requirements. Broadly speaking, $\mathcal{I}J$ for ABF and AIRSIF should be compared with $\mathcal{I}$ for UBF, J for PF, and JG for GIRF.

The computational effort allocated to each algorithm in Table S-1 is given in core minutes. UBF, ABF and ABF-IR parallelize readily, which is less true for PF and GIRF. Therefore, the UBF, ABF and ABF-IR implementations run on all available cores (40 for this experiment) whereas the PF and GIRF implementations run on a single core. If sufficient replications are being carried out to utilize all available cores, comparison of core minute utilization is equivalent to comparison of total computation time. However, a single replication of UBF, ABF or ABF-IR proceeds more quickly due to the parallelization.

ABF-IR and GIRF have computational time scaling quadratically with U in this example, whereas the other methods scale linearly. This is because the number of intermediate steps used, S, grows linearly with U.

The main purpose of this example is not to provide a comparison between the functional capabilities of the methods on interesting scientific problems. It is a toy example without the complexities that the methods are intended to address. This simple example does show clearly the quick decline of PF and the slower declines of GIRF and ABF as dimension increases.

	UBF	ABF	ABF-IR	GIRF	PF	BPF	EnKF
particles, J	—	100	50	1000	10000	5000	5000
bootstrap replications, $\mathcal{I}$	5000	500	200	—	—	—	—
guide simulations, G	—	—	—	50	—	—	—
lookahead lag, L	—	—	—	2	—	—	—
intermediate steps, S	—	—	$U/2$	U	—	—	—
neighborhood, $B_{u,n}$ or block size	{(u-1,n),(u-2,n), (u,n-1),(u,n-2)}			—	—	2	—
forecast mean, $\boldsymbol{\mu}(\boldsymbol{x}, s, t)$	—	—	$\boldsymbol{x}$		—	—	—
measurement mean, $h_{u,n}(x)$	—	—	x		—	—	x
$\tau = \overleftarrow{\mathrm{v}}_{u,n}(V, x)$	—	—	$\sqrt{V}$		—	—	—
$V = \overrightarrow{\mathrm{v}}_{u,n}(\tau, x)$	—	—	τ^2		—	—	τ^2
effort (core mins, $U = 32$)	0.1	1.1	1.1	0.1	0.0	0.0	0.0
effort (core mins, $U = 16$)	0.1	1.6	1.8	0.3	0.0	0.0	0.0
effort (core mins, $U = 8$)	0.2	2.1	3.2	0.7	0.1	0.1	0.0
effort (core mins, $U = NA$)	NA	NA	NA	NA	NA	NA	NA
effort (core mins, $U = NA$)	NA	NA	NA	NA	NA	NA	NA

Table S-1: Algorithmic settings for the correlated Brownian motion numerical example. Computational effort is measured in core minutes for running one filter, corresponding to a point on Figure 1 in the main text. The time taken for computing a single point using the parallel UBF, ABF and ABF-IR implementations is the effort divided by the number of cores, here 40. The time taken for computing a single point using the single-core GIRF, PF, BPF and EnKF implementations is equal to the effort in core minutes.

parameter	value	unit	description
$\bar{\beta}$	1560.6	year^{-1}	mean contact rate
$\mu_{\bullet D}^{-1}$	50.0	year	mean duration in the population
μ_{EI}^{-1}	7.0	day	latent period
μ_{IR}^{-1}	7.0	day	infectious period
σ_{SE}	0.150	$\text{year}^{1/2}$	process noise
a	0.500	—	amplitude of seasonality
α	1	—	mixing exponent
τ	4	year	delay from birth to entry into susceptibles
ρ	0.5	—	reporting probability
ψ	0.15	—	reporting overdispersion
G	400	—	gravitation constant

Table S-2: Table of parameters for the spatiotemporal measles transmission model

S6 The measles example

Table S-2 gives the model parameter values and Table S-3 gives the algorithmic settings used for the filters. The times in Table S-3 give the total time required by each algorithm to calculate all its results for Figure 2 in the main text, using 40 cores. The expected forecast function $\mu(x, s, t)$ needed for ABF-IR and GIRF was computed using a numerical solution to the deterministic skeleton of the stochastic model, i.e, a system of ODEs with derivative matching the infinitesimal mean function of the stochastic dynamic model. In the specifications of $h_u(x)$, $\overleftarrow{\mathrm{v}}_u(V, x, \theta)$ and $\overrightarrow{\mathrm{v}}_u(\theta, x)$, the latent process value x contains a variable C giving the cumulative removed infections in the current observation interval.

In Table S-3, we see that the effort allocated to UBF, ABF and PF scales linearly with U, since the number of bootstrap replications and particles is fixed in this experiment. GIRF computational effort scales fairly linearly in U, since its effort is dominated by the guide simulations (which are linear in U) rather than by the intermediate timestep calculations (which are quadratic in U since we carry out U intermediate calculations each of size linear in U). The effort allocated to ABF-IR scales with U^2, since ABF-IR is more parsimonious with guide simulations (all particles in one bootstrap replication share the same guide simulations) and so the intermediate timestep calculations dominate the effort. To obtain stable variance in the log likelihood estimate, the number of particles and bootstrap replications would have to grow with U. However, given a constraint on total computational resources, the number of particles and bootstrap replications would have to shrink as U increases. The limit studied in this experiment is a balance between the two: the assumption is that one is prepared to invest a growing amount of computational effort as the data grow, but this should not grow too fast. ABF-IR was permitted the greatest computational effort, but two considerations balance this.

1. Parallelization. UBF, ABF and ABF-IR are trivially parallelizable. The value of parallelization depends, among other things, on how many replications are being computed simultaneously and on how many cores are available. Nevertheless, it is helpful that the core minute effort requirement for ABF and ABF-IR can be divided by the number of available cores

	UBF	ABF	ABF-IR	GIRF	EnKF	PF	BPF
particles, J	1	100	50	1000	5000	10000	5000
replicates, $\mathcal{I}$	5000	500	200	—	—	—	—
guide simulations, G	—	—	—	50	—	—	—
lookahead lag, L	—	—	—	1	—	—	—
intermediate steps, S	—	—	$U/2$	U	—	—	—
neighborhood, $B_{u,n}$ or block size	$\{(u,n-1),(u,n-2)\}$			—	—	—	1
measurement mean, $h_{u,n}(x)$	—	—	ρC			—	—
$V = \overrightarrow{\mathrm{v}}_{u,n}(\psi,\rho,x)$	—	—	$\rho(1-\rho)C + \rho^2C^2\psi^2$			—	—
forecast mean, $\boldsymbol{\mu}(\boldsymbol{x},s,t)$	—	—	ODE model		—	—	—
$\psi = \overleftarrow{\mathrm{v}}_{u,n}(V,x)$	—	—	$\frac{\sqrt{V-\rho(1-\rho)C}}{\rho C}$		—	—	—
effort (core mins, $U=2$)	10.4	5.1	3.1	1.8	0.2	0.4	0.2
effort (core mins, $U=4$)	14.6	8.3	8.0	3.8	0.3	0.7	0.5
effort (core mins, $U=8$)	21.5	14.5	25.5	8.4	0.7	1.3	0.9
effort (core mins, $U=16$)	35.3	26.4	98.5	21.4	1.3	2.4	1.7
effort (core mins, $U=$)							

Table S-3: Algorithmic settings for the measles example calculations in Figures 2 and 3. Computational effort is measured in core minutes for running one filter, corresponding to a point on Figure 2. The time taken for computing a single point using the parallel UBF, ABF and ABF-IR implementations is the effort divided by the number of cores, here 40. The time taken for computing a single point using the single core GIRF, EnKF, PF and BPF implementations is equal to the effort in core minutes.

to give the computational time. Parallelizations of GIRF and PF can be constructed (Park and Ionides, 2020) but these involve non-trivial interaction between processors leading to additional algorithmic complexity and computational overhead.

2. Memory. The intermediate timestep calculations in ABF-IR and GIRF do not add to the memory requirement, and the memory demands of UBF, ABF and ABF-IR are distributed across the parallel computations. A basic PF implementation for a large model can become constrained by its memory requirement (linear in the number of particles) before it can match the processor effort employed by the other algorithms.

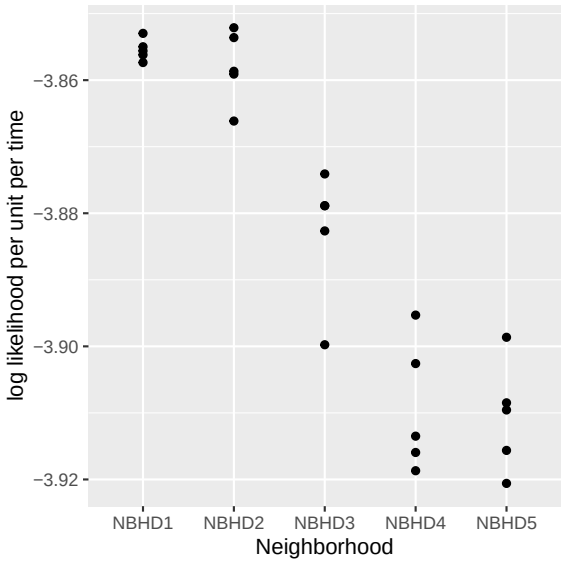

Figure S-2: log likelihood estimates for simulated data from the measles model using ABF, with varying neighborhoods.

S7 Varying the neighborhood for measles

We compared five different neighborhoods for the measles model:

NBHD1	One co-located lag	$\{(u, n-1)\}$
NBHD2	Two co-located lags	$\{(u, n-1), (u, n-2)\}$
NBHD3	Three co-located lags	$\{(u, n-1), (u, n-2), (u, n-3)\}$
NBHD4	Two co-located lags and the previous city	$\{(u, n-1), (u, n-2), (u-1, n)\}$
NBHD5	Two co-located lags and London	$\{(u, n-1), (u, n-2), (1, n)\}$

We filtered simulated data for $U = 40$ and $N = 104$, with 5 replications. We used ABF with 100 particles on each of 500 bootstrap replications. The results are shown in Fig. S-2. Larger neighborhoods should increase the expected likelihood, but their increased Monte Carlo variability can decrease the expected log likelihood due to Jensen's inequality. In this case, we see that a neighborhood of two co-located lags provides a reasonable bias-variance tradeoff. The time taken for the above calculation was insensitive to the size of the neighborhood. The total run time for each neighborhood in Fig. S-2 was 8.3 mins for NBHD1, 8.4 mins for NBHD2, 8.1 mins for NBHD3, 8.2 mins for NBHD4, 8.1 mins for NBHD5.

S8 A Lorenz-96 example

Our primary motivation for ABF and ABF-IR is application to population dynamics arising in ecological and epidemiological models. Geophysical models provide an alternative situation involving spatiotemporal data analysis. We compare methods on the Lorenz-96 model, a nonlinear chaotic system providing a toy model for global atmospheric circulation (Lorenz, 1996; van Kekem and Sterk, 2018). We consider a stochastic Lorenz-96 model with added Gaussian process noise (Park and Ionides, 2020) defined as the solution to the following system of stochastic differential equations,

$$dX_u(t) = \big\{\big(X_{u+1}(t) - X_{u-2}(t)\big) \cdot X_{u-1}(t) - X_u(t) + F\big\}dt + \sigma_p dB_u(t), \qquad u \in 1{:}U. \tag{S56}$$

We define $X_0 = X_U$, $X_{-1} = X_{U-1}$, and $X_{U+1} = X_1$ so that the U spatial locations are placed on a circle. The terms $\{B_u(t), u \in 1{:}U\}$ denote U independent standard Brownian motions. F is a forcing constant, and we use the value $F = 8$ which was demonstrated by Lorenz (1996) to induce chaotic behavior. The process noise parameter is set to $\sigma_p = 1$. The system is started with initial state $X_u(0)$ drawn as an independent normal random variable with mean 5 and standard deviation 2 for $u \in 1{:}U$. This initialization leads to short transient behavior. Observations are independently made for each dimension at $t_n = n$ for $n \in 1{:}N$ with Gaussian measurement noise of mean zero and standard deviation $\tau = 1$,

$$Y_{u,n} = X_u(t_n) + \eta_{u,n} \qquad\qquad \eta_{u,n} \sim N(0, \tau^2). \tag{S57}$$

We used an Euler-Maruyama method for numerical approximation of the sample paths of $\{\boldsymbol{X}(t)\}$, with timestep of 0.005. A simulation from this model is shown in Fig. S-4.

The ensemble Kalman filter (EnKF) is a widely used filtering method in weather forecasting for high dimensional systems (Evensen and van Leeuwen, 1996). EnKF involves a local Gaussian approximation which is problematic in highly nonlinear systems (Ades and Van Leeuwen, 2015). Methods that make local Gaussian assumptions like EnKF are necessary to scale up to the dimensions of the problems in weather forecasting. Figure S-3 shows that for a small number of units, the basic particle filter (PF) and GIRF out-perform EnKF. Then, as the number of spatial units increases, the performance of PF rapidly deteriorates whereas GIRF continues to perform well up to a moderate number of units. UBF, ABF, and particularly ABF-IR, scales well despite underperforming EnKF on this example. The additive Gaussian observation and process noise in the Lorenz-96 model is well suited to the approximations involved in EnKF. By contrast, it is less clear how to apply EnKF to discrete population non-Gaussian models such as the measles example, and how effective the resulting approximations might be.

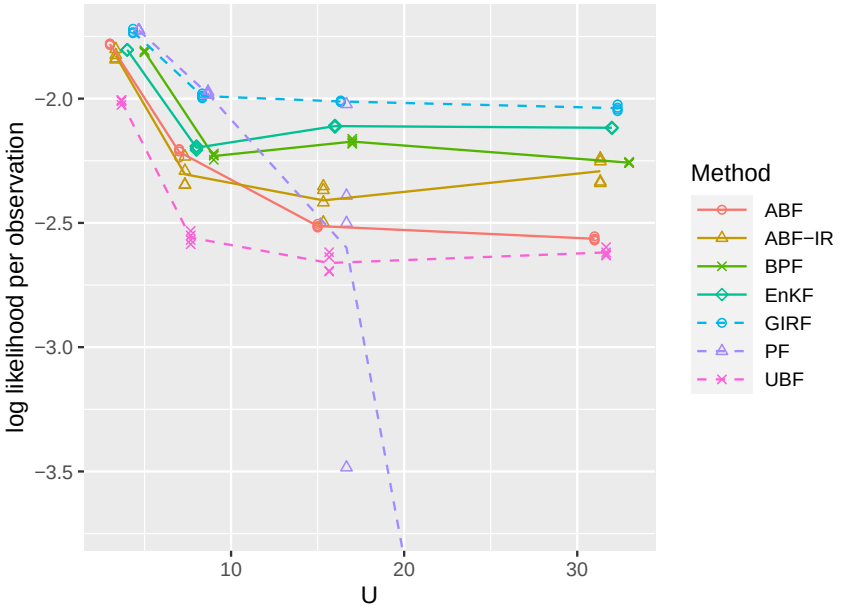

Figure S-3: Log likelihood estimates for a Lorenz-96 model of various dimensions. UBF, ABF and ABF-IR are compared with a guided intermediate resampling filter (GIRF), a standard particle filter (PF), a block particle filter (BPF) and an ensemble Kalman filter (EnKF).

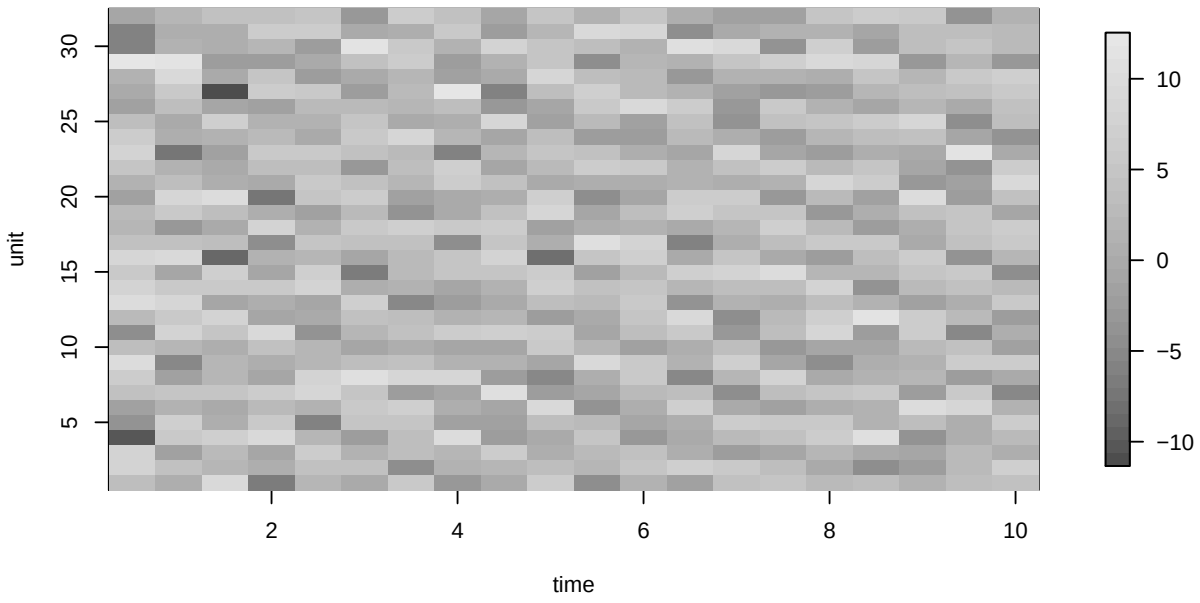

Figure S-4: Lorenz '96 simulation used in the main text

	UBF	ABF	ABF-IR	GIRF	PF	EnKF	BPF
particles, J	1	100	50	1000	50000	10000	10000
bootstrap replicates, $\mathcal{I}$	10000	500	50	—	—	—	—
guide simulations, G	—	—	—	50	—	—	—
lookahead lag, L	—	—	—	2	—	—	—
intermediate steps, S	—	—	$U/2$	U	—	—	—
neighborhood, $B_{u,n}$ or block size	$\{$(u,n-1),(u,n-2), (u-1,n),(u-2,n)$\}$			—	—	—	2
forecast mean, $\boldsymbol{\mu}(\boldsymbol{x}, s, t)$	—	—	ODE model		—	—	—
measurement mean, $h_{u,n}(x)$	—	—	x		—	x	—
$\tau = \overleftarrow{\mathrm{v}}_{u,n}(V, x)$	—	—	$\sqrt{V}$		—	—	—
$V = \overrightarrow{\mathrm{v}}_{u,n}(\tau, x)$	—	—	τ^2		—	τ^2	—
effort (core mins, $U = 4$)	8.2	1.9	0.9	1.0	0.5	0.1	0.1
effort (core mins, $U = 8$)	11.3	3.0	1.3	2.0	1.0	0.2	0.3
effort (core mins, $U = 32$)	28.5	9.0	4.7	9.1	3.9	0.8	0.9
effort (core mins, $U = NA$)	NA	NA	NA	NA	NA	NA	NA
effort (core mins, $U = NA$)	NA	NA	NA	NA	NA	NA	NA

Table S-4: Algorithmic settings for the Lorenz-96 numerical example. Computational effort is measured in core minutes for running one filter, corresponding to a point on Figure Figure S-3. The time taken for computing a single point using the parallel UBF, ABF and ABF-IR implementations is the effort divided by the number of cores, here 40. The time taken for computing a single point using the single-core GIRF, PF, EnKF and BPF implementations is equal to the effort in core minutes.

S9 A memory-efficient representation of ABF

The ABF pseudocode in the main text emphasizes the logical structure of the mathematical quantities computed, rather than describing a specific implementation. Arguably, an algorithm should specify not only what is to be computed, but also details of what variables should be created and saved to carry out these computations efficiently, and how computations and storage are shared across multiple locations when the algorithm is parallelized. We use the term algorithm to denote a higher level description of the quantities to be calculated and we will say that alternative pseudocodes arriving at the same quantities are representations of the algorithm. Here, we present an alternative representations of ABF which we call ABF_2, and we refer to the representation in the main text as ABF_1. The most concrete form of an algorithm is the actual computer code implementing the algorithm in a programming language. The implementation of ABF in `spatPomp`, used for numerical results in this paper, uses an embarrassingly parallel approach based on the representation ABF_1. This strategy facilitates robust and simple parallelization, and is appropriate when memory constraints are not limiting. To develop ABF_2, we set up notation similar to that used for the mathematical theory. Let

$$\gamma_{u,n,i,k} = \frac{1}{J}\sum_{j=1}^{J} \prod_{\tilde{u}:(\tilde{u},n-k)\in B_{u,n}} w^M_{\tilde{u},n-k,i,j} \tag{S58}$$

Also, let

$$\gamma^+_{u,n,i,0} = \frac{1}{J}\sum_{j=1}^{J} \prod_{(\tilde{u},n)\in B^+_{u,n}} w^M_{\tilde{u},n,i,j} \tag{S59}$$

Now set K to be the largest value of k for which $B^{[n-k]}_{u,n}$ is nontrivial for some (u,n), i.e., K is the largest temporal lag in any neighborhood. With this notation, we can write (S32) as

$$\begin{aligned}
\gamma^{MC,i}_{B_{u,n}} &= \prod_{k=0}^{K} \gamma_{u,n,i,k} \\
\gamma^{MC,i}_{B^+_{u,n}} &= \gamma^+_{u,n,i,0} \prod_{k=1}^{K} \gamma_{u,n,i,k}
\end{aligned}$$

This motivates the following representation of ABF.

ABF_2. Adapted bagged filter, representation 2.

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

For n in $1\!:\!N$

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

  Measurement weights: $w^{M}_{u,n,i,j} = f_{Y_{u,n}|X_{u,n}}\big(y^{*}_{u,n} \,|\, X^{\mathrm{P}}_{u,n,i,j}\big)$

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

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

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

  Compute $\gamma_{u,n+k,i,k}$ for k in $0\!:\!K$,

  Compute $\gamma^{+}_{u,n,i,0}$

End for

$$\ell^{\mathrm{MC}}_{u,n} = \log\left(\frac{\sum_{i=1}^{\mathcal{I}} \gamma^{+}_{u,n,i,0} \prod_{k=1}^{K} \gamma_{u,n,i,k}}{\sum_{i=1}^{\mathcal{I}} \prod_{k=0}^{K} \gamma_{u,n,i,k}}\right)$$

It is clearer from the ABF_2 representation than from the ABF_1 representation that we do not have to save every individual particle and its weight $w^{M}_{u,n,i,j}$ after the quantities $\boldsymbol{X}^{A}_{n,i}$, $\gamma_{u,n+k,i,k}$ and $\gamma^{+}_{u,n,i,0}$ have been computed. ABF_2 has an embarrassingly parallel implementation, since the replicates do not need to interact until they are combined to compute $\ell^{\mathrm{MC}}_{u,n}$. An embarrassingly parallel implementation therefore requires $O\big(U(KN+J)\big)$ memory for each replicate, and $O\big(UN\mathcal{I}K\big)$ memory when the results from each replicate are collected together.

By using additional communication, for example when the implementation is designed for a single core or multiple cores with shared memory, it is possible to further reduce the memory requirement. At time n, we need to save only $\boldsymbol{X}^{A}_{n,i}$ and $\gamma_{u,n-K:n+K,i,k}$ to compute $\ell^{\mathrm{MC}}_{u,n}$ and all subsequent quantities. Computing $\boldsymbol{X}^{A}_{n,i}$ requires $O(UJ)$ storage. Therefore, ABF can be implemented with memory requirement $O\big(U(K\mathcal{I}+J)\big)$, independent of N.

Supplementary 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.

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.

Liu, J. S. (2001). *Monte Carlo Strategies in Scientific Computing*. Springer, New York.

Lorenz, E. N. (1996). Predictability: A problem partly solved. *Proceedings of the Seminar on Predictability*, 1:1–18.

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.

van Kekem, D. L. and Sterk, A. E. (2018). Travelling waves and their bifurcations in the Lorenz-96 model. *Physica D: Nonlinear Phenomena*, 367:38–60.