Extensions to "Bagged filters for partially observed spatiotemporal systems"

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

September 6, 2020

This is a place to write additional material, not necessarily intended for publication, supplying comments or additional details or other related material.

Contents

E1 A bootstrap guide function within ASIF-IR

Brainstorming the naming of guide functions, it seems to me that the previous mean/variance based guide function could be called a *simulated moments* guide, and the *quantile guide function* is a variation on something that could be called a *bootstrap guide function* which arises most naturally when all quantiles are considered in the Park and Ionides (2020) quantile guide function.

In the context of ASIF-IR, here is some pseudocode. There are two possibilities given for constructing the bootstrap guide residuals. One is to calculate the residuals at time t_n and scale them according to $\sqrt{\frac{t_n - t_{n,s}}{t_n - t_{n,0}}}$, as done in the quantile guide function. Another, given here as [ALT], is to save residuals at each intermediate time point and use these. That could have advantages if the variability in the latent process changes considerably over one observation interval. For measles growing after a trough, there might perhaps be substantially more cases, and correspondingly more variability due to log scale noise, in the second half of a 2-week observation interval. This requires more memory, but not more calls to rprocess or flow, and if the mememory is available it might take comparable time. It may require more care to code efficiently, so perhaps [ALT] can wait.

ABF-IR. Adapted bagged filter with intermediate resampling. With updated guide function

---

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

For n in $1{:}N$

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

Guide residuals: $\boldsymbol{\epsilon}_{n,r,j} = \boldsymbol{X}^{G}_{n,r,j} - \boldsymbol{\mu}\big(\boldsymbol{X}^{\mathrm{A}}_{n-1,r}, t_{n-1}, t_n\big)$

[ALT] Guide residuals: $\boldsymbol{\epsilon}_{n,s,r,j} = \boldsymbol{X}^{G}_{n,s,r,j} - \boldsymbol{\mu}\big(\boldsymbol{X}^{\mathrm{A}}_{n-1,r}, t_{n-1}, t_{n,s}\big)$

$g^{\mathrm{R}}_{n,0,r,j} = 1 \quad$ and $\quad \boldsymbol{X}^{\mathrm{IR}}_{n,0,r,j} = \boldsymbol{X}^{\mathrm{A}}_{n-1,r}$

For s in $1{:}S$

Intermediate proposals: $\boldsymbol{X}^{\mathrm{IP}}_{n,s,r,j} \sim f_{\boldsymbol{X}_{n,s}|\boldsymbol{X}_{n,s-1}}\big(\,\cdot\,|\boldsymbol{X}^{\mathrm{IR}}_{n,s-1,r,j}\big)$

$\boldsymbol{\mu}^{\mathrm{IP}}_{n,s,r,j} = \boldsymbol{\mu}\big(\boldsymbol{X}^{\mathrm{IP}}_{n,s,r,j}, t_{n,s}, t_n\big)$

$g_{n,s,r,j} = \frac{1}{J^U}\prod_{u=1}^{U}\sum_{j=1}^{J} f_{Y_{u,n}|X_{u,n}}\Big(y^*_{u,n} \,|\, \mu^{\mathrm{IP}}_{u,n,s,r,j} + \sqrt{\frac{t_n - t_{n,s}}{t_n - t_{n,0}}}\boldsymbol{\epsilon}_{n,r,j}\,;\theta\Big)$

[ALT] $g_{n,s,r,j} = \frac{1}{J^U}\prod_{u=1}^{U}\sum_{j=1}^{J} f_{Y_{u,n}|X_{u,n}}\Big(y^*_{u,n} \,|\, \mu^{\mathrm{IP}}_{u,n,s,r,j} + \boldsymbol{\epsilon}_{n,S,r,j} - \boldsymbol{\epsilon}_{n,s,r,j}\,;\theta\Big)$

Guide weights: $w^{G}_{n,s,r,j} = g_{n,s,r,j}/g^{\mathrm{R}}_{n,s-1,r,j}$

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

$\boldsymbol{X}^{\mathrm{IR}}_{n,s,r,j} = \boldsymbol{X}^{\mathrm{IP}}_{n,s,r,r(r,j)} \quad$ and $\quad g^{\mathrm{R}}_{n,s,r,j} = g_{n,s,r,r(r,j)}$

End For

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

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

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

End for

$$\ell^{\mathrm{MC}}_{u,n} = \log\left(\frac{\sum_{r=1}^{\mathcal{R}}\sum_{j=1}^{J} w^{M}_{u,n,r,j} w^{P}_{u,n,r,j}}{\sum_{r=1}^{\mathcal{R}}\sum_{j=1}^{J} w^{P}_{u,n,r,j}}\right)$$

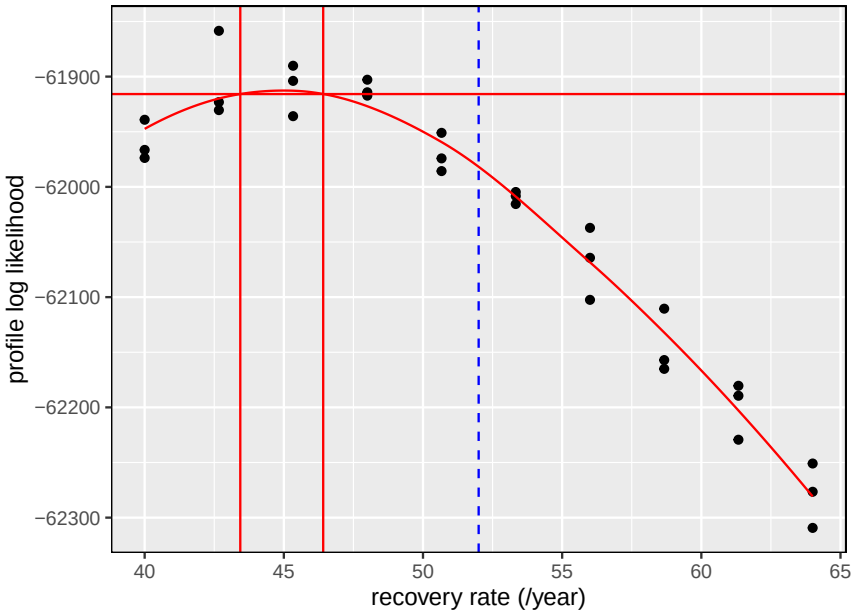

Figure E-1: Likelihood slice of the recovery rate parameter, computed via ABF on the measles model with $U = 40$ cities. The true parameter value for the simulation is marked with a blue dashed line.

E2 A likelihood slice across recovery rate for the measles model

As discussed in the main text, we expect parameters of within-city transmission that are shared between cities to be well identified. Figure **??** shows a likelihood slice for the per-capita recovery rate, μ_{IR} (in units of year^{-1}) on simulated data from the coupled model for $U = 40$ cities. As expected, the likelihood has peak in the vicinity of the truth at $\mu_{IR} = 52.0$ year^{-1}. The rapid decrease in the log likelihood away from this maximum dwarfs the Monte Carlo error in computing this log likelihood.

Fig E-1 took 6.2 minutes to evaluate each point using a parallelized implementation on a computing node with 36 cores. By contrast, Fig E-2 took 93.5 minutes, and Fig E-3 took 30.4 minutes. No doubt, all the implementations could be sped up with addition attention to the code.

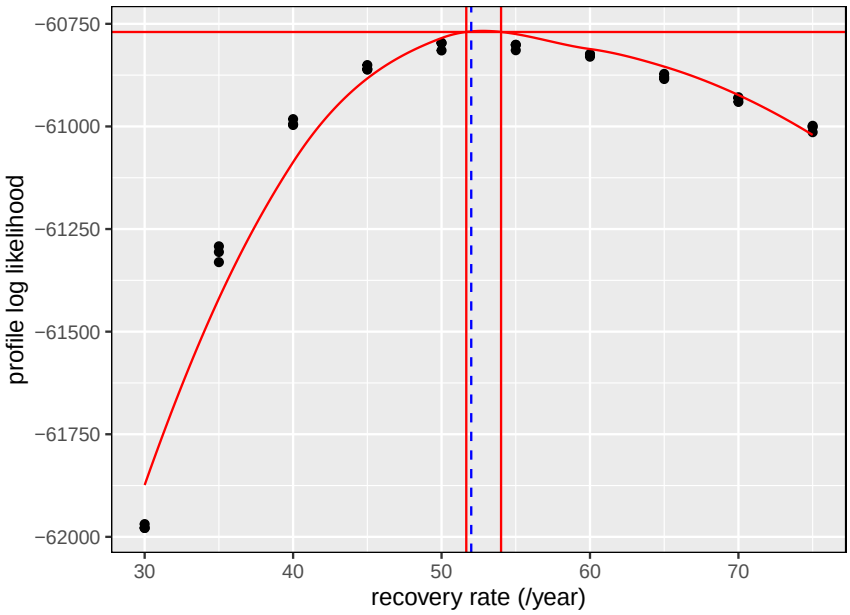

Figure E-2: Likelihood slice of the recovery rate parameter, computed via BPF on the measles model with $U = 40$ cities. The true parameter value for the simulation is marked with a blue dashed line.

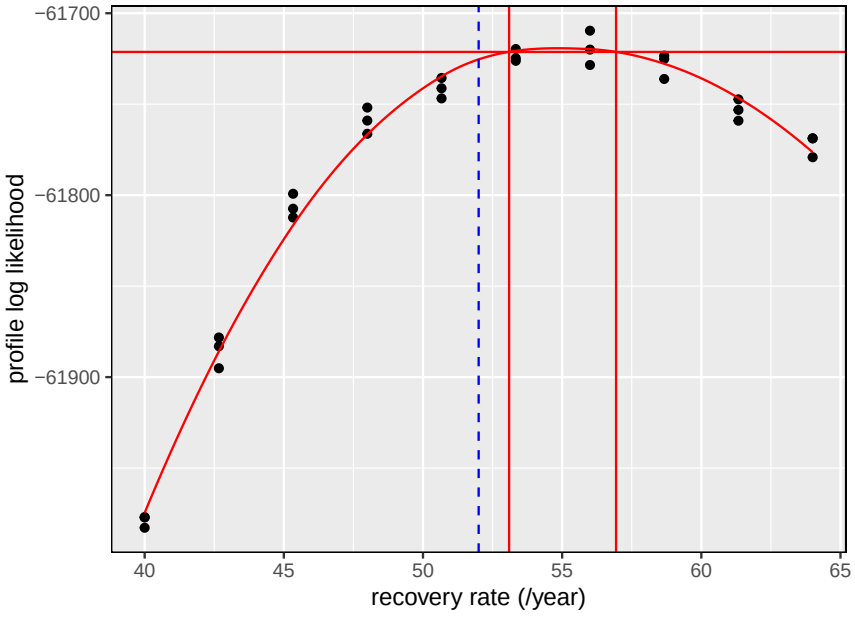

Figure E-3: Likelihood slice of the recovery rate parameter, computed via UBF on the measles model with $U = 40$ cities. The true parameter value for the simulation is marked with a blue dashed line.

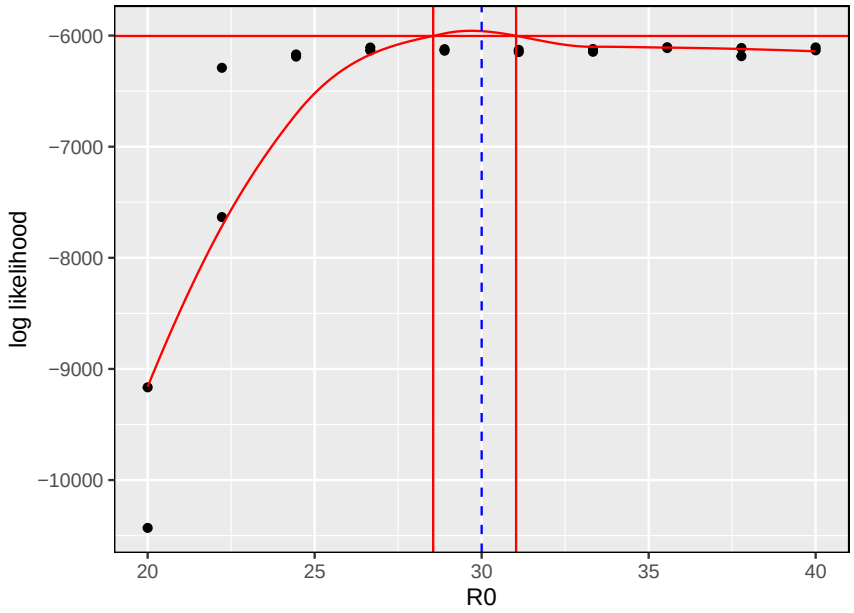

Figure E-4: Likelihood slice varying the coupling parameter, computed via ABF with $U = 40$ cities. The solid lines construct a 95% Monte Carlo adjusted confidence interval (Ionides et al., 2017). For this model realization, this interval includes the true parameter value identified by a blue dashed line.

E3 Slices over R_0

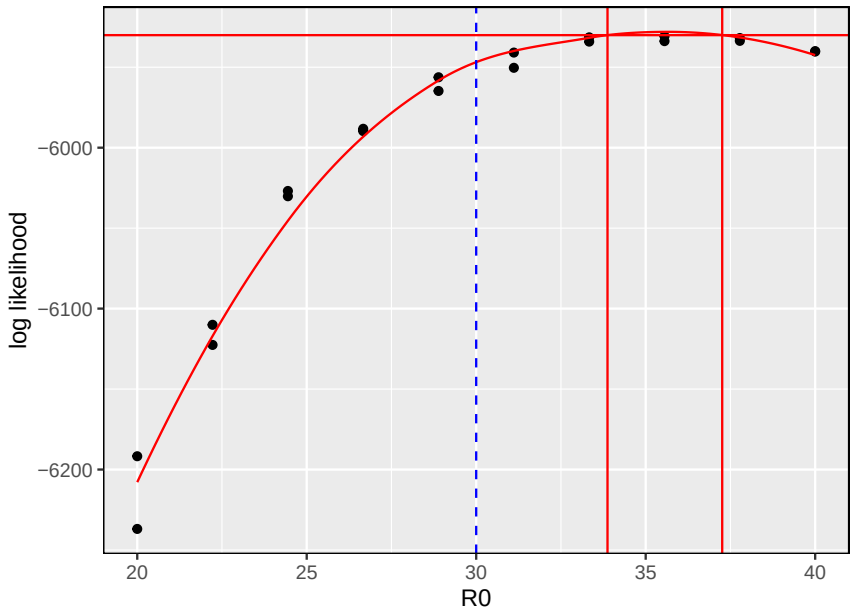

Figure E-5: Likelihood slice varying the coupling parameter, computed via BPF with $U = 40$ cities. The solid lines construct a 95% Monte Carlo adjusted confidence interval (Ionides et al., 2017). For this model realization, this interval includes the true parameter value identified by a blue dashed line.

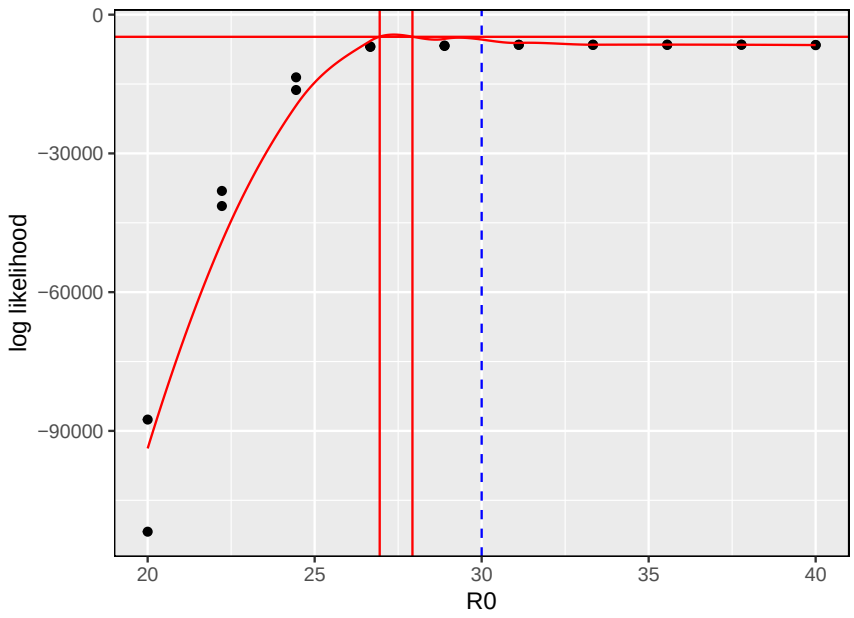

Figure E-6: Likelihood slice varying the coupling parameter, computed via UBF with $U = 40$ cities. The solid lines construct a 95% Monte Carlo adjusted confidence interval (Ionides et al., 2017). For this model realization, this interval includes the true parameter value identified by a blue dashed line.

E4 Comments on assumptions

1. The same ϵ is used in various different bounds. In principle, these could all be different bounds. However, it is conceptually simpler to have just a single small quantity. Also, the proof shows that these errors all have additive consequences for the overall error, so they may as well be on comparable scales.

2. Recall the following assumptions:

   Assumption A1. *There is an $\epsilon_{\mathrm{A1}} > 0$ 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, n, U, N, any bounded real-valued function $|h(x)| \leq 1$, and any value of $x_{B^c_{u,n}}$,*

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

   Assumption A4. *There exists $\epsilon_{\mathrm{A4}} > 0$ 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 $|C_{u,n}| \leq c$.

   Assumption B1. *There is an $\epsilon_{\mathrm{B1}} > 0$ 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, n, U, 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 $f_B(x_B) = f_{Y_B|X_B}(y^*_B|x_B)$,*

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

   Assumption B4. *Define $g_{X_A|X_B}$ via (**??**). Suppose there is an ϵ_{B4} 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*

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

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

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

The specification of $B^{2c}_{u,n}$ in Assumptions A4 and **??** does not have to be built using the same sets $B_{u,n}$ from Assumptions A1 and B1. It may be simpler to imagine that one set of neighborhoods carries out both functions, and if different neighborhoods are specified then the union of these neighborhoods provides one set satisfying both conditions.

E5 Deriving Assumption B5 from GIRF results

Recall the statement:

Assumption B5. *Let h be a bounded function with $|h(x)| \leq 1$. Let $\boldsymbol{X}^{\mathrm{IR}}_{n,S,j,r}$ be the Monte Carlo quantity constructed in ABF-IR, conditional on $\boldsymbol{X}^{\mathrm{A}}_{n-1,S,r} = \boldsymbol{x}^{\mathrm{A}}_{n-1,S,r}$. There is a constant $\mathcal{C}_0(U,N,S)$ such that, for all $\epsilon_{\mathrm{B6}} > 0$ and $\boldsymbol{x}^{\mathrm{A}}_{n-1,S,r}$, whenever the number of particles satisfies $J > \mathcal{C}_0(U,N,S)/\epsilon^3_{\mathrm{B6}}$,*

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

We check this is derived from the following result.

Theorem 3 (Park and Ionides, 2020). *Suppose multinomial resampling is used in GIRF, and that regularity conditions specify $C_1 > 0$, $C_2 > 0$, $0 < \rho < 1$. We don't need the exact conditions to check that the conclusion implies Assumption B5. If h is a measurable function such that $\|h\|_\infty \leq 1$ and $a > 1$ is an arbitrary constant, then we have*

$$\left| \frac{1}{J}\sum_{j=1}^{J} h(X^{F,j}_{t_N}) - \mathbb{E}\big[h(X_{t_N}) | Y_{1:N} = y_{1:N}\big] \right| \leq \frac{4aC_2(C_1+1)}{\rho\sqrt{J}}(NS+1) \tag{E1}$$

with probability at least

$$1 - \frac{(2NS+1)(NS+1)}{a^2}, \tag{E2}$$

given that $\sqrt{J} \geq 8\rho^{-2}aC_2(C_1+1)NS$.

In our context, $N = 1$ since we are carrying out guided sampling over one timestep. Now, set

$$\frac{(2NS+1)(NS+1)}{a^2} < \epsilon/2, \tag{E3}$$

say,

$$a = \sqrt{\frac{2(2S+1)(S+1)}{\epsilon}}. \tag{E4}$$

We want J large enough that

$$\frac{4aC_2(C_1+1)}{\rho\sqrt{J}}(NS+1) < \epsilon/2, \tag{E5}$$

since then the bounds in (E1) and (E2), together with the bound on h, give the bound in expectation in Assumption B5. Putting (E4) into (E5) gives

$$\frac{4C_2(C_1+1)\sqrt{2(2S+1)(S+1)}}{\rho\sqrt{\epsilon J}}(S+1) < \epsilon/2. \tag{E6}$$

and so

$$J > \frac{128(2S+1)(S+1)^3 C_2^2 (C_1+1)^2}{\epsilon^3 \rho^2} \tag{E7}$$

which implies that we can pick

$$C_0 = 128(2S+1)(S+1)^3 C_2^2 (C_1+1)^2 \rho^{-2} \tag{E8}$$

in Assumption B5.

E6 Some adapted simulation identities

The following proposition derives some results relating the adapted process to the filtering, prediction and likelihood evaluation tasks. The first two results, (E9) and (E10), are simpler identities for temporal dependence ignoring sequential calculations over space. They are provided primarily to give simpler analogs to the subsequent results.

Proposition E1. *Let* $g(\vec{x}) = f_{\vec{X}_0}\left[\prod_{n=1}^N f_{\vec{X}_n|\vec{X}_{n-1},\vec{Y}_n}\right]$ *be the joint density of* $\vec{X} = \vec{X}_{0:N}$ *for the adapted process, and let* $\gamma_B(\vec{x}) = \prod_{\tilde{n}=1}^N f_{Y_{B^{[\tilde{n}]}}|\vec{X}_{\tilde{n}-1}}$ *be the proper adapted weight on* B*. Then,*

$$f_{\vec{X}_n,\vec{Y}_{1:n}} = \int f_{\vec{X}_0}\left[\prod_{m=1}^n f_{\vec{X}_m|\vec{X}_{m-1},\vec{Y_m}} f_{\vec{Y}_m|\vec{X}_{m-1}}\right] d\vec{x}_{0:n-1} \tag{E9}$$

$$\begin{aligned} f_{\vec{X}_{n+1}|\vec{Y}_{1:n}} &= \frac{\int f_{\vec{X}_0}\left[\prod_{m=1}^n f_{\vec{X}_m|\vec{X}_{m-1},\vec{Y_m}} f_{\vec{Y}_m|\vec{X}_{m-1}}\right] f_{\vec{X}_{n+1}|\vec{X}_n}\, d\vec{x}_{0:n}}{\int f_{\vec{X}_0}\left[\prod_{m=1}^n f_{\vec{X}_m|\vec{X}_{m-1},\vec{Y_m}} f_{\vec{Y}_m|\vec{X}_{m-1}}\right] d\vec{x}_{0:n}} \\ &= \frac{\mathbb{E}_g\left[\gamma_{1:U\times\{1:n\}}(\vec{X})\, f_{\vec{X}_{n+1}|\vec{X}_n}(\vec{x}_{n+1}\,|\,\vec{X}_n)\right]}{\mathbb{E}_g\left[\gamma_{1:U\times\{1:n\}}(\vec{X})\right]} \end{aligned} \tag{E10}$$

$$f_{\vec{X}_n,Y_{A_{u,n}}} = \int f_{\vec{X}_0}\left[\prod_{m=1}^n f_{\vec{X}_m|\vec{X}_{m-1},\vec{Y_m}} f_{Y_{A_{u,n}^{[m]}}|\vec{X}_{m-1}}\right] d\vec{x}_{0:n-1} \tag{E11}$$

$$\begin{aligned} f_{Y_{A_{u,n}}} &= \int f_{\vec{X}_0}\left[\prod_{m=1}^n f_{\vec{X}_m|\vec{X}_{m-1},\vec{Y_m}} f_{Y_{A_{u,n}^{[m]}}|\vec{X}_{m-1}}\right] d\vec{x}_{0:n} \\ &= \mathbb{E}_g\left[\gamma_{A_{u,n}}(\vec{X})\right] \end{aligned} \tag{E12}$$

$$\begin{aligned} f_{X_{u:U,n}|Y_{A_{u,n}}} &= \frac{\int f_{\vec{X}_0}\left[\prod_{m=1}^n f_{\vec{X}_m|\vec{X}_{m-1},\vec{Y_m}} f_{Y_{A_{u,n}^{[m]}}|\vec{X}_{m-1}}\right] dx_{A_{u,n}}}{\int f_{\vec{X}_0}\left[\prod_{m=1}^n f_{\vec{X}_m|\vec{X}_{m-1},\vec{Y_m}} f_{Y_{A_{u,n}^{[m]}}|\vec{X}_{m-1}}\right] d\vec{x}_{0:n}} \\ &= \frac{\mathbb{E}_g\left[\gamma_{A_{u,n}}(\vec{X})\, f_{X_{u:U,n}|X_{A_{u,n}}}(x_{u:U,n}|X_{A_{u,n}})\right]}{\mathbb{E}_g\left[\gamma_{A_{u,n}}(\vec{X})\right]} \end{aligned} \tag{E13}$$

Proof. Suppose inductively that (E9) holds for some n. Using the Markov property,

$$f_{\vec{Y}_{1:n+1},\vec{X}_{n+1}} = \int f_{\vec{Y}_{1:n},\vec{X}_n} f_{\vec{Y}_{n+1},\vec{X}_{n+1}|\vec{X}_n} d\vec{x}_n \tag{E14}$$

Substituting (E9) into (E14) and checking the case $n = 1$ shows that (E9) holds also for $n+1$, and therefore (E9) holds for all $n \geq 1$ by induction. The first equality in (E10) follows from (E9) by standard manipulation of condition densities. The second equality in (E10) employs the definition $\gamma_{1:U\times\{1:n\}} = \gamma_{1:U\times\{1:n\}}(\vec{X}_{0:N}) = \prod_{m=1}^{n} f_{\vec{Y}_m|\vec{X}_{m-1}}$. (E11) is analogous to (E9). Integrating (E11) over $\vec{x}_n$ gives the joint likelihood expression in (E12). Integrating only over $x_{1:u-1,n}$ and conditioning gives (E13). □

E7 Variations on local weights

These are some ideas that were written up before we settled on ABF/ABF-IR as the main focus for this manuscript.

E7.1 Moving from weak weights to proper local weights

Consider a bivariate situation where we draw $(X_j^{(1)}, Y_j^{(1)}) \sim f_{XY}^{(1)}(x, y)$ and resample with weights $g_X(x)g_Y(y)$ to obtain $(X_j^{(2)}, Y_j^{(2)})$ targeting

$$f_{XY}^{(2)}(x, y) = \frac{g_X(x)g_Y(y)f_{XY}^{(1)}(x, y)}{\int g_X(x')g_Y(y')f_{XY}^{(1)}(x', y')\, dx'dy'}. \tag{E15}$$

Now suppose we want to use $\{(X_j^{(2)}, Y_j^{(2)}), j \in 1:J\}$ to obtain a sample from the marginal X-distribution of

$$f_{XY}^{(3)}(x, y) = \frac{h_X(x)h_Y(y)f_{XY}^{(1)}(x, y)}{\int h_X(x')h_Y(y')f_{XY}^{(1)}(x', y')\, dx'dy'}. \tag{E16}$$

It appears we should resample $X_{1:J}^{(2)}$ with weights

$$W_j = \frac{h_X(X_j^{(2)})}{g_X(X_j^{(2)})}.$$

We should do an analogous calculation in the IIF algorithm. Note that the local reweighting (here, local to the X-component) is, in fact, a proper reweighting even though we ignore the Y-component. This follows, despite the dependence in f_{XY} because of the product structure of the weights $g_X(x)g_Y(y)$ and $h_X(x)h_Y(y)$.

E7.2 Using the weakly weighted fixed lag smoothing distribution as a proposal

For the time being, we suppose we are working with a single generic island r, and we suppress the identity of this island from the notation. Write $B_{u,n}^{+} = B_{u,n} \cup (u, n)$ and $B_{u,n}^{-} = B_{u,n} \cap \{(\tilde{u}, \tilde{n}) : \tilde{n} < n\}$. We need a proposal for $X_{B_{u,n}^{+}}$ that can be locally reweighted at (u, n) to solve the proper local spatiotemporal filtering problem. Let a weakly smoothed particle approximation $X_{[u,n],\tilde{u},\tilde{n},1:J}^{WS}$ be defined as follows, for $(\tilde{u}, \tilde{n}) \in B_{u,n}^{+}$.

$$X_{[u,n],\tilde{u},\tilde{n},j}^{WS} = \begin{cases} X_{\tilde{u},\tilde{n},\langle n,\tilde{n},j\rangle}^{WF} & \text{for } \tilde{n} < n \\ X_{\tilde{u},\tilde{n},j}^{WP} & \text{for } \tilde{n} = n \end{cases} \tag{E17}$$

This collection of particles, by construction, approximates a weakly weighted smoothed density,

$$f_{X_{B^+_{u,n}}|\vec{Y}_{1:n-1}}(x_{B^+_{u,n}} \,|\, \vec{y}^*_{1:n-1}\,;\sigma_{1:U}\Omega_{1:U}). \tag{E18}$$

We can appeal to the structure of the relaxed measurement process to give

$$f_{X_{B^-_{u,n}}|\vec{Y}_{B^-_{u,n}}}(x_{B^-_{u,n}} \,|\, y^*_{B^-_{u,n}}\,;\sigma_{\tilde{u}}) \tag{E19}$$

$$\propto \quad f_{X_{B^-_{u,n}}|Y_{1:n-1}}(x_{B^-_{u,n}} \,|\, y^*_{1:n-1}\,;\sigma_{\tilde{u}}\Omega_{\tilde{u}}) \prod_{(\tilde{u},\tilde{n})\in B^-_{u,n}} \frac{f_{Y_{\tilde{u},\tilde{n}}|X_{\tilde{u},\tilde{n}}}(y^*_{\tilde{u},\tilde{n}}|x_{\tilde{u},\tilde{n}}\,;\sigma_{1:U})}{f_{Y_{\tilde{u},\tilde{n}}|X_{\tilde{u},\tilde{n}}}(y^*_{\tilde{u},\tilde{n}}|x_{\tilde{u},\tilde{n}}\,;\sigma_{1:U}\Omega_{1:U})}. \tag{E20}$$

Now, $X^{WS}_{[u,n],B^+_{u,n},j}$, considered as a particle taking values in $\mathbb{X}^{B^+_{u,n}}$ and intended to represent the target density $f_{X_{B^+_{u,n}}|Y_{B_{u,n}}}$, has proper weight given by

$$w^P_{u,n,j} = \prod_{(\tilde{u},\tilde{n})\in B^-_{u,n}} \frac{f_{Y_{\tilde{u},\tilde{n}}|X_{\tilde{u},\tilde{n}}}\big(y^*_{\tilde{u},\tilde{n}}|X^{WS}_{[u,n],\tilde{u},\tilde{n},j}\,;\sigma_u\big)}{f_{Y_{\tilde{u},\tilde{n}}|X_{\tilde{u},\tilde{n}}}\big(y^*_{\tilde{u},\tilde{n}}|X^{WS}_{[u,n],\tilde{u},\tilde{n},j}\,;\sigma_u\Omega_u\big)} \prod_{(\tilde{u},\tilde{n})\in B^\sharp_{u,n}} f_{Y_{\tilde{u},\tilde{n}}|X_{\tilde{u},\tilde{n}}}\big(y^*_{\tilde{u},\tilde{n}}|X^{WS}_{[u,n],\tilde{u},\tilde{n},j}\,;\sigma_u\big). \tag{E21}$$

The (u,n) marginal of these particles therefore represents

$$f_{X_{u,n}|Y_{B_{u,n}}}(x_{u,n} \,|\, y^*_{B_{u,n}}\,;\sigma_{1:U}) \approx f_{X_{u,n}|Y_{A_{u,n}}}(x_{u,n} \,|\, y^*_{A_{u,n}}\,;\sigma_{1:U}). \tag{E22}$$

Therefore,

$$\sum_{j=1}^{J}\left\{\frac{w^P_{u,n,j}}{\sum_{k=1}^J w^P_{u,n,k}}\right\} f_{Y_{u,n}|X_{u,n}}(y^*_{u,n} \,|\, X^{WS}_{[u,n],u,n,j}\,;\sigma_u) \approx f_{Y_{u,n}|Y_{A_{u,n}}}(y^*_{u,n} \,|\, y^*_{A_{u,n}}\,;\sigma_{1:U}). \tag{E23}$$

It might seem from (E23) that we need to build particles solving a different fixed lag smoothing problem at each spatiotemporal location (u,n). However, we see in the IIF algorithm that these particles are constructed as weighted collections of a single collection of weakly filtered particles and their ancestors defined over all space and time. A simple computational approach is to carry out the entire weak filtering, saving all the particles and the identity of their parent, and then carry out the proper local weighting computations. In principle, one can carry out the weak filtering and reweighting synchronously, enabling memory saving by discarding weakly filtered particles once they no longer show up in any current or future spatiotemporal neighborhood $B_{u,n}$. Doing this requires additional attention to communication in a multi-processor implementation of IIF. So long as storage is not prohibitive (say, 1000 particles at 100 locations for 1000 time points, each with an 8-byte double precision real value, requiring 800Mb storage) it is tempting to treat the weak filtering and reweighting as separate tasks, solved sequentially.

The importance sampling in (E21) occurs in dimension $\big|B_{u,n}\big|$. If this dimension is large, the weights may be numerically stable. The curse of dimensionality is addressed since this dimension is assumed not to depend on U or N.

There is a tradeoff as the weakly filtered distribution approaches the target distribution, the weights in (E21) become more balanced but the weak filtering itself becomes more problematic.

E7.3 Weak weights versus other weak filters

The core idea of the island filter is that a collection of weaker filters, each of which is numerically feasible despite the dimensionality, can be combined in a neighborhood of each spatiotemporal location.

An alternative approach to weak filtering is to carry out local filtering, where each island focuses on making one draw from the conditional distribution of $X_{1:U,n}$ given $X_{1:U,n-1}$ and $Y_{1:U,n}$. A particle filter using $f_{X_{1:U,n}|Y_{1:U,n},X_{1:U,n-1}}$ as its proposal is said to be *fully adapted*, with some intuition that this might be the ideal (usually inaccessible) proposal choice given information available at time n, even though it is not necessarily better than a basic unadapted proposal (Johansen and Doucet, 2008). Our weak adapted filter (WAP) will circumvent some issues for the COD by using the collection of particles only to make a sample from $f_{X_{1:U,n}|Y_{1:U,n},X_{1:U,n-1}}$ rather than representing the whole distribution. This is an easier problem since it requires only comparison of nearby particles, which will necessarily have similar weights (assuming the measurement density is uniformly continuous) and will therefore be less susceptible to the unbalanced weights causing the COD for importance sampling. To obtain the full filtering distribution by reweighting these draws, the proper weight is $f_{Y_{1:U,n}|X_{1:U,n-1}}$ which could itself be hard to compute for large U. However, in the context of IIF, we need to compute these weights only locally.

In practice, the WAP can be combined with the weak weights. An advantage of weak weights is that one can see how, at least in principle, allowing islands to ignore all but a fixed number of observations can lead to algorithms formally beating the COD.

It may be difficult to see how a one-good-particle GIRF can have local properties beating the COD. However, a filter such as GIRF with some good properties should be able to be used with weak weights to improve scaling.

A continuous time approximation may be helpful. In continuous time, supposing that massive amounts of information never show up instantaneously, the curse of too much information can be avoided. This enables importance sampling to give Monte Carlo simulation of $f_{\vec{X}(t+\delta)|\vec{Y}(t+\delta),\vec{X}(t)}$ and evaluation of $f_{\vec{Y}(t+\delta)|\vec{X}(t)}$ for sufficiently small δ. However, we may need small δ, and hence a large number of importance sampling steps, as U increases. In general, to avoid COD, one expects $\delta \propto 1/U$, as studied in the context of GIRF by Park and Ionides (2020). This produces a problem for SMC if we look for a proper global reweighting of the islands. One then expects the Monte Carlo variance of the log weights for each island to grow linearly with $1/U$. Put another way, the Monte Carlo variance in the importance sampling estimate of the likelihood should grow linearly with the information available. We are saved by the local reweighting. Specifically, let $1:N$ be the discretization of the continuous system, so $t_n = n\delta$. Then, we aim to sample from $f_{X_{u,n}|Y_{B_{u,n}}}$ (i.e., the localized version of the filtering problem) by taking the (u,n) component from an adapted filter proposal, combined with an unadapted prediction step (we can do this, since the prediction step is constructed by the filter particles before they are weighted and resampled). We write $B_{u,n}^{[m]} = \{(\tilde{u},\tilde{n}) \in B_{u,n} : \tilde{n} = m\}$. Then, our proposal density for $X_{B_{u,n}}^{+}$ is the marginal distribution on $\mathbb{X}_{B_{u,n}}^{+}$ of a joint density on $\mathbb{X}_{A_{u,n}}^{+}$ given by

$$f_{X_{n,1:u}|\vec{X}_{n-1}}(x_{n,1:u} \,|\, \vec{x}_{n-1}) \prod_{\tilde{n}=1}^{n-1} f_{\vec{X}_{\tilde{n}}|\vec{Y}_{\tilde{n}},\vec{X}_{\tilde{n}-1}}(\vec{x}_{\tilde{n}} \,|\, \vec{y}^{*}_{\tilde{n}}, \vec{x}_{\tilde{n}-1}), \tag{E24}$$

Note that the proposal for $X_{n,u}$ doesn't condition on the data $y^{*}_{n,1:u-1}$ at time n but preceding in the spatial ording. This differs from a regular particle filter applied to the spatiotemporal

ordering. The density of a proposal for $X^+_{B_{u,n}}$ from (E24) involves integrating out the proposal density for $X_{B^c_{u,n}}$. However, since we are interested in the distribution at (u,n), by the weak coupling assumption we incur only a small error by instead setting $x_{B^c_{u,n}}$ to any arbitrary value. It is natural to choose for this purpose the value of $X_{B^c_{u,n}}$ for the filter particles, a choice which is convenient and approximates the omitted integral. The resulting weight to target a draw from the filter density, $f_{X_{u,n}|Y_{A_{u,n}}}(x_{u,n} \,|\, y^*_{A_{u,n}})$, via a draw $X^P_{A_{u,n}}$ from (E24) is

$$\frac{f_{X_{n,1:u},Y_{n,1:u}|\vec{X}_{n-1}}(x_{n,1:u}, y^*_{n,1:u} \,|\, \vec{x}_{n-1}) \prod_{m=1}^{n-1} f_{X_{B^{[m]}_{u,n}},Y_{B^{[m]}_{u,n}}|\vec{X}_{m-1}}\Big(X^P_{B^{[m]}_{u,n}}, y^*_{B^{[m]}_{u,n}} \,\Big|\, \vec{X}^P_{m-1}\Big)}{f_{X_{n,1:u}|\vec{X}_{n-1}}(x_{n,1:u} \,|\, \vec{x}_{n-1}) \prod_{m=1}^{n-1} f_{X_{B^{[m]}_{u,n}}|Y_{B^{[m]}_{u,n}},\vec{X}_{m-1}}\Big(X^P_{B^{[m]}_{u,n}} \,\Big|\, y^*_{B^{[m]}_{u,n}}, \vec{X}^P_{m-1}\Big)} \tag{E25}$$

$$= f_{Y_{1:u-1,n}|X_{1:u-1,n}}(y^*_{1:u-1,n} \,|\, X^P_{1:u-1,n}) \prod_{m=1}^{n-1} f_{Y_{B^{[m]}_{u,n}}|X_{B^{[m]}_{u,n}}}\Big(y^*_{B^{[m]}_{u,n}} \,\Big|\, X^P_{B^{[m]}_{u,n}}\Big) \tag{E26}$$

E8 Comments on the guide function

The plug-and-play guide function of Park and Ionides (2020) involves an additive moment-based approximation which is suitable if the process and measurement models are close to Gaussian. Here, we explore this assumption in the case of the measles model.

The guide function uses an approximation to $\mathrm{Var}(Y_n|X_{n-1,s} = x_{n-1,s})$ where $X_{n-1,s} = X\big(t_{n-1,s}\big)$ is the latent process at intermediate time $t_{n-1} = t_{n-1,0} < t_{n-1,s} < t_{n-1,S} = t_n$ for $1 \le s < S$. Then, the predictive likelihood at t_n given $X_{n-1,s} = x_{n-1,s}$ is approximated using the measurement model with center at an approximation $\hat{x_n} \approx \mathbb{E}\big[Y_n \,|\, X_{n-1,s} = x_{n-1,s}\big]$ and variance approximated by

$$\mathrm{Var}[Y_n \,|\, X_n = \hat{x_n}\big] + \mathrm{Var}\big[h(X_n) \,|\, X_{n-1,s} = x_{n-1,s}\big] \tag{E27}$$

where $h(x) = \mathbb{E}[Y_n|X_n = x]$. We use (E27) in place of

$$\mathrm{Var}[Y_n \,|\, X_{n-1,s} = x_{n-1,s}\big] \tag{E28}$$

since (E28) is hard to calculate.

For a toy analog to measles, consider the situation where $X = e^U$ and $Y = Xe^V = e^{U+V}$ where $U \sim \mathcal{N}(\sigma^2)$ and $V \sim \mathcal{N}(\tau^2)$. We have rescaled cases and reports for this analogy, but that is unimportant here. The approximation (E27) becomes

$$\mathrm{Var}[Y_n \,|\, X_n = \hat{x_n}\big] + \mathrm{Var}\big[h(X_n) \,|\, X_{n-1,s} = x_{n-1,s}\big] = \big(e^{2\sigma^2} - e^{\sigma^2}\big) + \big(e^{2\tau^2} - e^{\tau^2}\big), \tag{E29}$$

with the exact calculation being

$$\mathrm{Var}[Y_n \,|\, X_{n-1,s} = x_{n-1,s}\big] = \big(e^{2(\sigma^2+\tau^2)} - e^{(\sigma^2+\tau^2)}\big). \tag{E30}$$

The ratio of (E29) to (E30) in the case $\sigma = \tau$ is

$$\frac{2}{e^{2\sigma^2} + e^{\sigma^2}}. \tag{E31}$$

This approaches 1 for σ small, since in this case the measurement model is essentially additive. However, the ratio rapidly decays as σ heads past 1.

```
> f <- function(sigma) 2/(exp(sigma^2)+exp(2*sigma^2))
> sigma <- (0:10)/10
> cbind(sigma,f=f(sigma))
      sigma   f
 [1,]   0.0 1.0000000
 [2,]   0.1 0.9850996
 [3,]   0.2 0.9415762
 [4,]   0.3 0.8728320
 [5,]   0.4 0.7841173
 [6,]   0.5 0.6819546
 [7,]   0.6 0.5734335
 [8,]   0.7 0.4654657
 [9,]   0.8 0.3640918
[10,]   0.9 0.2739351
[11,]   1.0 0.1978760
```

In the measles context, rescaling to match this toy example, we consider $\tau^2 = \psi^2 = 0.116$ and $\sigma^2 = \sigma_{\mathrm{SE}}^2 \times 14 = 0.0878^2 \times 14 = 0.108$. This gives a ratio of

```
> sigma <- sqrt(0.108) ; tau <- 0.116
>  ( exp(2*sigma^2)+exp(2*tau^2)-exp(sigma^2)-exp(tau^2) ) /
+   ( exp(2*(sigma^2+tau^2)) -  exp(sigma^2+tau^2) )
[1] 0.9654933
>
```

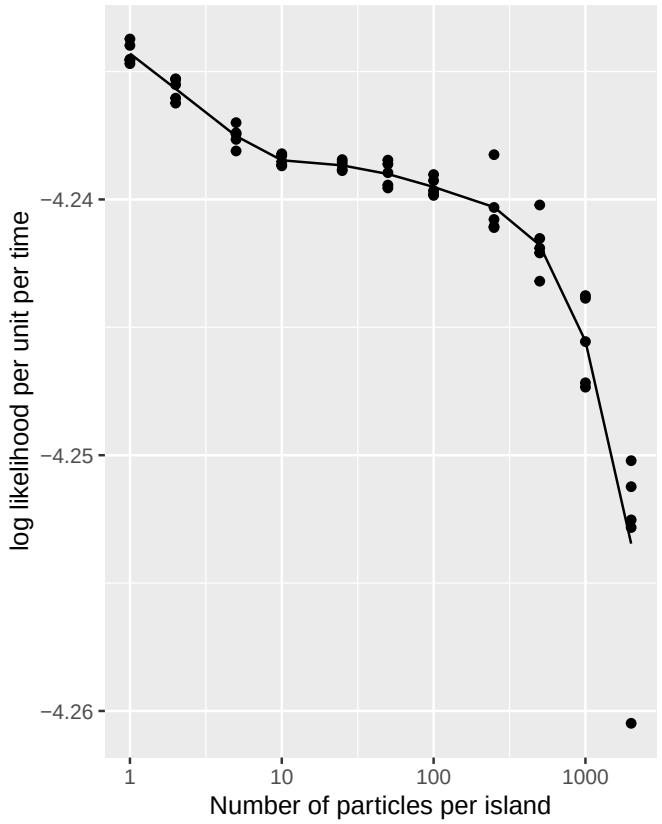

Figure E-7: log likelihood estimates for simulated data from the measles model using ABF, with varying algorithmic parameters.

E9 The islands vs particle per island trade-off for measles

We investigated the tradeoff between the number of islands, $\mathcal{R}$, and the number of particles per island, J, for the measles model. We filtered simulated data for $U = 40$ and $N = 104$, with 5 replications. The results are shown in Figure E-7. The total time taken was 166.7 mins for ABF1, 165.1 mins for ABF2, 168.8 mins for ABF3, 179.1 mins for ABF4, 214.0 mins for ABF5, 274.7 mins for ABF6. The algorithmic settings were as follows:

J	$\mathcal{R}$	time
2000	500	166.7
1000	1000	165.1
500	2000	168.8
250	4000	179.1
100	10000	214.0
50	20000	274.7
25	40000	398.6
10	40000	304.4
5	40000	269.4
2	40000	252.4
1	40000	241.1

E10 Varying parameters for the Lorenz example

This is to assess the hypothesis that scalable particle filter methods may have larger advantages over EnKF when the nonlinear dynamics are not enveloped in relatively large amounts of noise. First, we tried reducing the process noise to $\sigma_p = 0.25$. Reducing process noise while maintaining measurement noise appears to makes it harder for guiding and adapted simulation to operate. However, the relatively large amount of Gaussian measurement noise fits the assumptions of EnKF.

The Jan 2020 version of the manuscript used $\sigma_p = 0.25$ but we later reverted to $\sigma_p = 1$.

The simulated data are plotted in Figure **??**, and the algorithmic parameters with corresponding run times are listed in Table **??**.

The numerical results here were not kept updated, and have been deleted.

E11 Choosing the observation interval

This is to assess the hypothesis that scalable particle filter methods may have larger advantages over EnKF when the time between observations is larger. the nonlinear dynamics are not enveloped in relatively large amounts of noise. We set the process noise to $\sigma_p = 1$, the measurement noise to $\tau = 1$ and observation times $t_n = n$ instead of $t_n = n/2$ as was the case previously.

The numerical results here were not kept updated with spatPomp changes, and have been deleted. The hypothesis seemed to be refuted. The system is already highly nonliner on the scale $t_n = n/2$ and quite challenging for particle methods. Making the problem harder can favor EnKF since its failure mode is much softer.

E12 Varying intermediate steps for ABF-IR on Lorenz

This is to investigate the effect of changing the number of intermediate steps, S, in ABF-IR for the Lorenz '96 model.

The numerical results here were not kept updated with spatPomp changes, and have been deleted.

Parameter values were $\sigma_p = 0.25$ and $\tau = 0.25$. We found improvement as S increased, with diminishing returns. It looked like $S = U/2$ was sufficient - for example, $S = 5$ was as good as $S = 10$ when $U = 10$.

Changing to $\tau = 1$, keeping $\sigma_p = 0.25$, gave dramatically different results. The results favored EnKF, and the improvement with increasing S was small, in which case ABF might be preferred to ABF-IR. An interpretation is that the reduced measurement noise is both problematic for an unguided method and helpful for the guided method. The combination of small σ_p and high τ is problematic for particle methods in a highly nonlinear, chaotic system: particles that find themselves in the wrong place cannot catch up by favorable choice of process noise.

E13 Measles scaling with reduced noise

The simpler ABF method performed better than the more complex ABF-IR and GIRF algorithms on the measles example of Section **??**. This may be because the coupling is weak, or because the guide functions used by ABF-IR and GIRF are not guiding well. To investigate further, we repeated the experiment with reduced process noise, setting $\sigma_{SE} = 0.02$. We hypothesized that reducing the process noise would improve the predictive capability of the deterministic skeleton $\mu^P_{u,n,s,\ell,j}$ and therefore improve the guide function.

The numerical results here were not kept updated with spatPomp changes, and have been deleted. Due to changes in the measles model, it is not clear if the previous results have much value. We found that, in this example, ABF and ABF-IR both performed apparently successfully and equally.

E14 Bagged local filters

In this section, we present a method that is related to, but different from, UBF or ABF. The method runs a collection of independent particle filters, each of which is based on a spatially local sequence of observations. Each of the independent islands, $i \in \mathcal{I}$, is a particle filter using the joint propagation kernel,

$$f_{\mathbf{X}_n|\mathbf{X}_{n-1}},$$

and the local observation density for a certain spatial unit $\mathfrak{u}(i)$ in $1:U$,

$$f_{Y_{\mathfrak{u}(i),n}|X_{\mathfrak{u}(i),n}}(y_{\mathfrak{u}(i),n}|\cdot).$$

The propagated and filter particles at time n and location u in the i-th island will be denoted by $X^P_{u,n,i,j}$ and $X^F_{u,n,i,j}$. We write an estimate of $f_{Y_{\mathfrak{u}(i),n}|Y_{\mathfrak{u}(i),1:n-1}}$ by the i-th island as

$$\hat{\ell}_{n,i} := \frac{1}{J}\sum_{j=1}^{J} f_{Y_{\mathfrak{u}(i),n}|X_{\mathfrak{u}(i),n}}(y_{\mathfrak{u}(i),n}|X^P_{\mathfrak{u}(i),n,i,j}).$$

We use the particles in $\mathcal{I}$ islands to estimate the likelihood $f_{Y_{1:U,1:N}}$ as follows. The approach is similar to that of ABF or ABF-IR, in that each particle is weighted by the prediction weight, and a weighted sum of measurement weights gives an estimate of $f_{Y_{u,n}|Y_{A_{u,n}}}$. The prediction weight $w^P_{u,n,i,j}$ is given by

$$w^P_{u,n,i,j} = \prod_{(\tilde{u},\tilde{n})\in B(u,n)} w^{u,n}_{\tilde{u},\tilde{n},i,j}$$

where

$$w^{u,n}_{\tilde{u},\tilde{n},i,j} = \begin{cases} \hat{\ell}_{\tilde{n},i} & \text{if } \mathfrak{u}(i) = \tilde{u} \\ f_{Y_{\tilde{u},\tilde{n}}|X_{\tilde{u},\tilde{n}}}(y_{\tilde{u},\tilde{n}}|X^P_{\tilde{u},\tilde{n},i,a^{u,n}_{\tilde{n}}(j;i)}) & \text{if } \mathfrak{u}(i) \neq \tilde{u}. \end{cases} \tag{E32}$$

Here, $a^{u,n}_{\tilde{n}}(j;i)$ is defined such that

$$X^P_{\tilde{n},i,a^{u,n}_{\tilde{n}}(j;i)} \text{ is the ancestor particle of } \begin{cases} X^F_{n,i,j} & \text{if } \mathfrak{u}(i) < u \\ X^P_{n,i,j} & \text{if } \mathfrak{u}(i) \geq u, \end{cases} \tag{E33}$$

where a particle is considered its own ancestor when $n = \tilde{n}$. We define $a^{u,n}_{\tilde{n}}(j;i)$ as in (E33) because the prediction weights when estimating $f_{Y_{u,n}|Y_{A_{u,n}}}$ cannot use particles that already contain the information about $Y_{\mathfrak{u}(i),n}$ with $\mathfrak{u}(i) \geq u$. According to a standard notation for ancestor particles, the ancestor particle of $X^F_{n,i,j}$ in the i-th island can be written as

$$X^F_{n',i,a^n_{n'}(j;i)}.$$

If $n = n'$, we let $a^n_{n'}(j;i) = j$. Using this notation, we can express $a^{u,n}_{\tilde{n}}(j;i)$ as

$$a^{u,n}_{\tilde{n}}(j;i) = \begin{cases} a^n_{\tilde{n}-1}(j;i) & \text{if } \mathfrak{u}(i) < u \\ a^{n-1}_{\tilde{n}-1}(j;i) & \text{if } \mathfrak{u}(i) \geq u. \end{cases}$$

The measurement weights are given by

$$w^M_{u,n,i,j} = \begin{cases} f_{Y_{u,n}|X_{u,n}}(y_{u,n}|X^F_{u,n,i,j}) & \text{if } \mathfrak{u}(i) < u \\ f_{Y_{u,n}|X_{u,n}}(y_{u,n}|X^P_{u,n,i,j}) & \text{if } \mathfrak{u}(i) \geq u. \end{cases}$$

The Monte Carlo estimate of $f_{Y_{u,n}|Y_{A_{u,n}}}$ is given by, as in ABF(-IR),

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

The pseudocode for this method (which is tentatively called a bagged local filter) is given below.

BLF. Bagged Local Filter

For each $i \in 1:\mathcal{I}$, run the bootstrap particle filter with the joint propagation kernel $f_{\mathbf{X}_n|\mathbf{X}_{n-1}}$ and local measurement density $f_{Y_{\mathfrak{u}(i),n}|X_{\mathfrak{u}(i),n}}(y_{\mathfrak{u}(i),n}|\cdot)$ for some spatial unit $\mathfrak{u}(i)$.

For n in $1:N$

$w_{u,n,i,j}^{P} = \prod_{(\tilde{u},\tilde{n})\in B(u,n)} w_{\tilde{u},\tilde{n},i,j}^{u,n}$ where $w_{\tilde{u},\tilde{n},i,j}^{u,n}$ is defined in (E32)

Measurement weights: $w_{u,n,i,j}^{M} = \begin{cases} f_{Y_{u,n}|X_{u,n}}(y_{u,n}|X_{u,n,i,j}^{F}) & \text{if } \mathfrak{u}(i) < u \\ f_{Y_{u,n}|X_{u,n}}(y_{u,n}|X_{u,n,i,j}^{P}) & \text{if } \mathfrak{u}(i) \geq u. \end{cases}$

End For

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

References

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.

Johansen, A. M. and Doucet, A. (2008). A note on the auxiliary particle filter. *Statistics and Probability Letters*, 78:1498–1504.

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.