Poisson Approximate Likelihood versus the block particle filter for a spatiotemporal measles model

Kunyang He, Yize Hao and Edward L. Ionides

University of Michigan, Ann Arbor

2025-07-05

Abstract

Filtering algorithms for high-dimensional nonlinear non-Gaussian partially observed stochastic processes provide access to the likelihood function and hence enable likelihood-based or Bayesian inference for this methodologically challenging class of models. A novel Poisson approximate likelihood (PAL) filter was introduced by Whitehouse et al. (2023). PAL employs a Poisson approximation to conditional densities, offering a fast approximation to the likelihood function for a certain subset of partially observed Markov process models. PAL was demonstrated on an epidemiological metapopulation model for measles, specifically, a spatiotemporal model for disease transmission within and between cities. At face value, Table 3 of Whitehouse et al. (2023) suggests that PAL considerably out-performs previous analysis as well as an ARMA benchmark model. We show that PAL does not outperform a block particle filter and that the lookahead component of PAL was implemented in a way that introduces substantial positive bias in the log-likelihood estimates. Therefore, the results of Table 3 of Whitehouse et al. (2023) do not accurately represent the true capabilities of PAL.

1 Introduction

Investigations of the metapopulation dynamics of measles (i.e., studying how measles infection moves within and between collections of spatially distinct populations) have motivated various methodological innovations for inference on high-dimensional partially observed stochastic processes (Xia et al., 2004; Park and Ionides, 2020; Ionides et al., 2023). The analysis by Whitehouse et al. (2023) (henceforth, WWR) provides a new approach to model-based inference on population dynamics via the Poisson approximate likelihood (PAL) filtering algorithm. WWR claimed impressive results on both a low-dimensional rotavirus transmission model and a high-dimensional measles model. On close inspection, the rotavirus results turned out to be overstated (Hao et al., 2024) leading to a published correction (Whitehouse et al., 2025). However, the spatiotemporal measles results were unaffected by that correction, and our present purpose is to revisit this example.

In Section 2 we show by direct numerical experimentation that the WWR implementation of PAL has substantial positive bias in the log-likelihood estimate, and so the use of log-likelihood to support the use of the method is flawed. While doing this, we show that a widely applicable block particle filter (BPF) is adequate on this problem. In Section 3 we explain theoretically how the positive bias for PAL in the measles example arises as a result of the lookahead mechanism included in the implementation of PAL that WWR used for this model. The lookahead mechanism was not used by WWR for the rotavirus analysis since a basic version of PAL was sufficient for that lower-dimensional example. Section 4 is a concluding discussion.

For our current purposes, we do not have to delve into the details of the measles data and model, so we provide only a brief overview. The data are measles case counts aggregated over 2-week intervals for forty of the largest towns in England and Wales, from 1949 to 1964. The data and the model are derived from Park and Ionides (2020), who built on a long tradition of models described therein. Recently, weekly data

for more towns have become publicly available (Korevaar et al., 2020), but we limit ourselves to the data used by WWR. The latent process model describes an integer count of infected, susceptible and recovered individuals in each town. The rate of disease transmission within cities follows widely used epidemiological equations. Transmission between pairs of cities follows a power law, diminishing with distance between the cities. This is known as a gravity model. Overdispersion for the latent dynamics is achieved by placing multiplicative gamma white noise on the transmission rate. The measurement model is a discretized Gaussian approximation to an overdispersed binomial (Park and Ionides, 2020) or Gaussian noise on a binomial rate (WWR). Park and Ionides (2020) used a particle filter known as a guided intermediate resampling filter (GIRF). Recently, BPF has been shown to have good performance on this class of models (Ionides et al., 2023, 2024; Ning and Ionides, 2023). Therefore, we compare PAL with BPF.

2   Numerical experiments for PAL and BPF

There are many possible numerical experiments that could be conducted to compare filters on spatiotemporal measles models. Here, we choose experiments to investigate two specific hypotheses:

H1 The lookahead version of the Poisson approximate likelihood estimator of WWR, which we call PAL_L, can have substantial positive bias on its log-likelihood estimate. This occurs in the spatiotemporal example of WWR.

H2 The bias in the PAL_L log-likelihood estimate scales approximately linearly with the number of spatial units.

We consider probabilistic filtering algorithms that are defined in the context of a model, its model parameters, and additional algorithmic parameters. Additionally, we require data, and this can either be the real historical measles data or can be simulated from another model that may or may not be the same model with the same parameters as used for the filter. We also have a choice of how many spatial units to include, with each unit being one UK town in the measles example. The experimental variables, and the list of values we consider for them, are summarized in Table 1 and further described below.

A central part of our reasoning is that a probabilistic forecasting filter (i.e., one that solves the one-step prediction problem without looking ahead to future data) cannot, on average, obtain higher log-likelihood than the exact prediction distribution, when the data are generated by the exact model. This is a restatement of the well-known fact that log-likelihood is a proper scoring rule (Gneiting and Raftery, 2007). To apply this property, we must work with simulated data so that the true generating model is known. In high dimensions, it is generally not possible to calculate the exact prediction distribution to within arbitrarily small error. Provably consistent Monte Carlo methods have intolerable Monte Carlo error, and that is the reason why approximation algorithms such as PAL_L are being invented. There are two special situations where we can establish accurately the true log-likelihood for the spatiotemporal measles models of interest: (i) when the number of spatial units is very small; (ii) when there is no spatial coupling, so the filtering problem can be solved independently for each single unit. In both these cases, a basic particle filter provides the desired, essentially exact, log-likelihood estimate. The basic particle filter is consistent and unbiased for the likelihood (Del Moral, 2004) and so, when its estimates have low empirical variance, it provides the required ground truth. In practice, order 10^5 particles give a highly accurate log-likelihood for one unit, but the strong sensitivity of the particle filter to the curse of dimensionality (Bengtsson et al., 2008) means that quantifiably exact estimates rapidly become unfeasible. Therefore, we consider two choices of size for the system, $U = 1$ and $U = 40$, with the latter being the size of the system tested by WWR. To allow for the study of systems without coupling (i.e., measles transmission between cities), we consider two model variations, C_1 and C_2, where C_1 is the original model with gravity coupling used by WWR, and C_2 is a modification where direct movement of infection between cities is replaced by a constant background rate of importation of infection.

One approach to address the limitations of particle filters for high-dimensional systems is to take advantage of the possibility to design the filter's prediction distribution for the nth observation, at time t_n, to build

Table 1: Variables for the numerical experiments and their set of values.

Variable	Description	Value 1	Value 2	Value 3
F	filter algorithm	PAL_L	PAL_V	BPF
J	number of particles	$J_1 = 5 \times 10^3$	$J_2 = 10^5$	
U	number of spatial units	$U_1 = 1$	$U_2 = 40$	
f_C	spatiotemporal coupling for filter	$C_1 = (g \neq 0, \iota = 0)$	$C_2 = (g = 0, \iota \neq 0)$	
f_{proc}	process model for filter	Euler	discrete	
f_{meas}	measurement model for filter	binomial	Gaussian	
f_θ	parameter for filter	$\hat{\theta}^*_{BPF}$	$\hat{\theta}_{PAL}$	
s_C	spatiotemporal coupling for simulation	$C_1 = (g \neq 0, \iota = 0)$	$C_2 = (g = 0, \iota \neq 0)$	
s_{proc}	process model for simulation	Euler	discrete	
s_{meas}	measurement model for simulation	binomial	Gaussian	
s_θ	parameter for simulation	$\hat{\theta}^*_{BPF}$	$\hat{\theta}_{PAL}$	

a prediction distribution that incorporates data occurring at, or subsequent to, time t_n. Such lookahead filters cannot be implemented for forecasting, but can be used for likelihood evaluation. This is the approach adopted by WWR's lookahead filter, PAL_L, which incorporates a particle filter to address overdispersion in the data.

There is no mathematical theorem prohibiting a lookahead filter estimating a higher likelihood than the truth, and in an extreme case the lookahead filter could just assert a one-step prediction distribution with all its mass on the actual data. Lookahead filters therefore need careful theoretical guarantees if we want to use a high likelihood estimate as evidence for both the success of the filter and (when doing data analysis) evidence supporting the model used to construct the filter. We will investigate the theory behind PAL_L later, in Section 3, but for now we just identify the potential hazard.

The experimental variables listed in Table 1 are now described in more detail:

F **The filtering algorithm**. PAL_L is the lookahead filter of WWR, and PAL_V is the plain, so-called vanilla, implementation. BPF is the block particle filter of Rebeschini and Van Handel (2015) implemented as `bpfilter` in spatPomp (Asfaw et al., 2024). Likelihood optimization for PAL_L and PAL_V is conducted using stochastic gradient descent and automatic differentiation, using the implementation by WWR. Likelihood optimization for BPF is conducted using the iterated BPF algorithm (Ning and Ionides, 2023; Ionides et al., 2024) implemented as `ibpf` in spatPomp. For $U = 1$, and for $U = 40$ with $g = 0$, BPF is identical to a basic particle filter. For simplicity, we use `bpfilter` from the spatPomp package even when `pfilter` from the pomp package is equivalent.

f_C **Spatiotemporal coupling for the filter model**. The choice $C = C_1$ corresponds to the coupling used by WWR, with spatial movement of infection ($g \neq 0$) and no background immigration of infection from outside the study system ($\iota = 0$). In order to test the methods on a high-dimensional system for which the true likelihood is known to a good degree of accuracy, we also consider setting C_2, without coupling ($g = 0$) and with compensating immigration to prevent permanent extinction of measles in small towns ($\iota \neq 0$). When $U = U_1 = 1$, we use the largest city, London, for which stochastic extinctions are very unlikely. Note that, when $U = 1$, the value of g becomes irrelevant.

f_{proc} **Latent process transition model for the filter**. For $f_{\mathrm{proc}} = $ discrete, a single gamma-distributed dynamic noise variable is chosen for each observation interval. This is the choice made by WWR. For $f_{\mathrm{proc}} = $ Euler, independent gamma noise variables are included in each Euler time step, so that the limit of the process model (as the Euler time step decreases) corresponds to a continuous-time over-dispersed Markov chain. Both $f_{\mathrm{proc}} = $ discrete and $f_{\mathrm{proc}} = $ Euler are implemented with a step of 1/2 week for the multinomial transitions conditional on the gamma noise.

f_{meas} **Measurement model for the filter**. The choice of WWR is $f_{\mathrm{meas}} = $ binomial, corresponding

Table 2: Combinations of variable values used for each experiment, E_k, $k = 1 \ldots 15$. The simulation settings, s_C, s_{proc}, s_{meas} and s_θ, are applicable only when we filter simulated data rather than real data.

E	F	J	U	f_C	f_{proc}	f_{meas}	f_θ	s_C	s_{proc}	s_{meas}	s_θ
E_1	BPF	J_2	U_1	C_1	Euler	Gaussian	$\hat{\theta}^*_{BPF}$	C_1	Euler	Gaussian	$\hat{\theta}^*_{BPF}$
E_2	PAL_V	J_1	U_1	C_1	discrete	binomial	$\hat{\theta}_{PAL}$	C_1	Euler	Gaussian	$\hat{\theta}^*_{BPF}$
E_3	PAL_V	J_2	U_1	C_1	discrete	binomial	$\hat{\theta}_{PAL}$	C_1	Euler	Gaussian	$\hat{\theta}^*_{BPF}$
E_4	PAL_L	J_1	U_1	C_1	discrete	binomial	$\hat{\theta}_{PAL}$	C_1	Euler	Gaussian	$\hat{\theta}^*_{BPF}$
E_5	PAL_L	J_2	U_1	C_1	discrete	binomial	$\hat{\theta}_{PAL}$	C_1	Euler	Gaussian	$\hat{\theta}^*_{BPF}$
E_6	PAL_L	J_1	U_1	C_1	discrete	binomial	$\hat{\theta}_{PAL}$	————	data	————	
E_7	BPF	J_1	U_1	C_1	Euler	Gaussian	$\hat{\theta}^*_{BPF}$	————	data	————	
E_8	BPF	J_2	U_2	C_2	Euler	Gaussian	$\hat{\theta}^*_{BPF}$	C_2	Euler	Gaussian	$\hat{\theta}^*_{BPF}$
E_9	PAL_V	J_1	U_2	C_1	discrete	binomial	$\hat{\theta}_{PAL}$	C_2	Euler	Gaussian	$\hat{\theta}^*_{BPF}$
E_{10}	PAL_L	J_1	U_2	C_1	discrete	binomial	$\hat{\theta}_{PAL}$	C_2	Euler	Gaussian	$\hat{\theta}^*_{BPF}$
E_{11}	BPF	J_2	U_2	C_1	Euler	Gaussian	$\hat{\theta}^*_{BPF}$	————	data	————	
E_{12}	PAL_V	J_1	U_2	C_1	discrete	binomial	$\hat{\theta}_{PAL}$	————	data	————	
E_{13}	PAL_V	J_2	U_2	C_1	discrete	binomial	$\hat{\theta}_{PAL}$	————	data	————	
E_{14}	PAL_L	J_1	U_2	C_1	discrete	binomial	$\hat{\theta}_{PAL}$	————	data	————	
E_{15}	PAL_L	J_2	U_2	C_1	discrete	binomial	$\hat{\theta}_{PAL}$	————	data	————	

to binomial measurements with truncated multiplicative Gaussian noise on the reporting rate, i.e., the expected fraction of infections that are reported. The basic PAL algorithm requires a binomial measurement model, but the SMC-PAL extension permits noise on the measurement probability. We also consider $f_{\text{meas}} = $ Gaussian, corresponding to a discretized Gaussian measurement model. This choice leads to a measurement model that can directly be evaluated, without costly Monte Carlo calculation, assisting with efficient Monte Carlo inference.

f_θ **Model parameter vector for the filter**. PAL_L and PAL_V are evaluated at optimized parameter vectors for each data set. For PAL, there is generally no true POMP model for which PAL is an exact filter. However, we give PAL a reasonable chance to show its capabilities by optimizing it using the code provided by WWR. When $U > 1$, we must decide which parameter are shared between units and which are unit-specific. For PAL_L and PAL_V, we used case A from WWR, for which parameters are shared between units. This ran much more quickly than their case C, and is sufficient to make our point. E_{12}-E_{15} use the parameter values published by WWR for their case A. For E_{11}, we used case C.

s_{proc} **Process model for the simulation**. Always set to $s_{\text{proc}} = $ Euler since simulations were carried out using an implementation of the model in spatPomp.

s_{meas} **Measurement model for the simulation**. Always set to $s_{\text{meas}} = $ Gaussian since simulations were carried out using an implementation of the model in spatPomp.

s_θ **Model parameter vector for the simulated data**. To have an essentially exact likelihood evaluation using BPF, we have $s_\theta = f_\theta$ for all BPF situations. The parameters were fixed at a value near the maximum likelihood estimate for the data.

We consider two software platforms for the experiments. All PAL_L and PAL_V calculations were carried out using the Python code provided by WWR, and all BPF calculations were carried out using the spatPomp R package. Simulations were carried out using spatPomp, since our reasoning depends critically on the particle filter calculations being carried out with the data drawn from the model assumed by the filter. We present results for a single simulation for the experimental treatments with simulated data. That decision simplifies the experimental design and permits the computational effort to focus on a few direct comparisons between the methods on a small number of simulated datasets. The experiments carried our are described in Table 2.

Table 3: Log-likelihood estimate, λ, for each experiment described in Table 2. Estimates derive from averaging 20 replications on a natural scale so that the basic particle filter estimate is unbiased. The standard error, σ, is a jack-knife estimate implemented via the logmeanexp function in the pomp R package. When $\sigma \gg 1$, this standard error is unreliable and we conclude only that the error is large. ARMA gives the log-likelihood for an autoregressive moving average benchmark, and NegBinom is an autoregressive negative binomial benchmark.

E	F	J	U	f_C	λ	σ	ARMA	NegBinom
E_1	BPF	J_2	U_1	C_1	-3133.06	0.03	-3190.08	-3248.91
E_2	PAL_V	J_1	U_1	C_1	-3208.66	0.56	-3190.08	-3248.91
E_3	PAL_V	J_2	U_1	C_1	-3206.26	0.22	-3190.08	-3248.91
E_4	PAL_L	J_1	U_1	C_1	-2984.61	0.92	-3190.08	-3248.91
E_5	PAL_L	J_2	U_1	C_1	-2993.40	0.45	-3190.08	-3248.91
E_6	PAL_L	J_1	U_1	C_1	-2449.86	1.86	-2561.32	-2751.23
E_7	BPF	J_1	U_1	C_1	-2501.15	0.32	-2561.32	-2751.23
E_8	BPF	J_2	U_2	C_2	-79611.19	0.18	-82301.84	-84818.27
E_9	PAL_V	J_1	U_2	C_1	-81445.26	7.78	-82301.84	-84818.27
E_{10}	PAL_L	J_1	U_2	C_1	-78223.41	12.86	-82301.84	-84818.27
E_{11}	BPF	J_2	U_2	C_1	-67761.82	0.78	-69168.10	-72401.20
E_{12}	PAL_V	J_1	U_2	C_1	-74709.92	206.77	-69168.10	-72401.20
E_{13}	PAL_V	J_2	U_2	C_1	-74137.62	4.23	-69168.10	-72401.20
E_{14}	PAL_L	J_1	U_2	C_1	-64014.93	60.48	-69168.10	-72401.20
E_{15}	PAL_L	J_2	U_2	C_1	-64848.76	2.25	-69168.10	-72401.20

Each experiment, E_k, has two primary outcomes, a log-likelihood estimate, λ_k, and its standard error, σ_k. These results are tabulated in Table 3, together with benchmark log-likelihoods for a log-ARMA(2,1) model and a negative binomial autoregressive model.

Here, experiments E_1–E_5 provide a computationally tractable comparison using simulated data on a single unit. E_1 provides a ground truth for this particular model, a single-city SEIR model. Comparing λ_2 and λ_3 with λ_1, we see that PAL_V, without the lookahead, performs as expected for an approximate filter. On this relatively easy task, it produces stable estimates, with log-likelihood values somewhat below the truth. However, both E_2 and E_3 failed to outperform the ARMA benchmark in our experiments, suggesting that PAL continues to face challenges when dealing with complex SEIR models and overdispersion. Since PAL_V is filtering using a model that differs slightly from the data generating model, we expect to see a small shortfall, with $\lambda_2 < \lambda_1$ and $\lambda_3 < \lambda_1$. The difference, $\lambda_2 - \lambda_3$, is statistically indistinguishable from zero in this experiment, showing that $J_1 = 5000$ particles is adequate for PAL_V on a single city.

E_4 and E_5 demonstrate the positive bias of PAL_L both at a usual number of particles and for an intensive calculation that may not be possible on larger problems. The best estimates of this bias are $\lambda_4 - \lambda_3$ and $\lambda_5 - \lambda_3$, since PAL_L and PAL_V target the same quantity in the limit as $J \to \infty$. Comparison of E_5 and E_3 shows that, even with a large number of particles and a low-dimensional dynamic model, PAL_L provides an over-stated log-likelihood estimate.

Comparing E_4 and E_5, we see that the smaller number of particles for E_4 leads to a higher log-likelihood estimate. This would not happen in a situation where the likelihood estimate is unbiased with finite variance: in that situation, a higher number of particles is expected to lead to a higher log-likelihood estimate due to reduced negative bias resulting from Jensen's inequality.

Experiments E_6 and E_7 introduce the actual data, while still restricting to a single spatial unit. We see that PAL_L reports a higher log-likelihood than BPF. At face value, this could be because the PAL approximation

is a superior model to the partially observed Markov process model implemented by BPF. Or, it could be because BPF suffers from heavy negative bias due to high Monte Carlo error combined with Jensen's inequality. The latter is not the case due to BPF's empirically low Monte Carlo error. We have just discovered in E_1–E_5 that PAL_L reports an over-stated log-likelihood when the truth is known, so the most plausible explanation of E_6 and E_7 is simply that the same phenomenon occurs on the data.

Experiments E_8–E_{15} investigate a 40 unit system. For E_8, E_9 and E_{10}, we simulated from a model with the coupling parameter between towns set to zero. That was done to study a situation where a block particle filter gives a consistent and low-variance estimate of the exact log-likelihood, calculated as λ_8. We see the same story as the single-unit case, where the positive bias of PAL_L is estimated by $(\lambda_{10} - \lambda_9)/40 = 80.55$ per unit. This bias is large enough that we also obtain $\lambda_{10} > \lambda_8$, with the difference being $(\lambda_{10} - \lambda_8)/40 = 34.69$ per unit. We see that the $U = 40$ results scale approximately linearly compared to $U = 1$, providing us with supporting evidence for our hypothesis H2.

Experiments E_{11}–E_{15} consider coupled models for all 40 cities in the full, real dataset. Here, E_{11} uses the same choice of shared and unit-specific parameters as model C in Table 3 of WWR, re-optimized using iterated BPF to account for differences between this model and the model of WWR. Experiments E_{12}–E_{15} directly use the parameters provided by WWR for their model A. The choice of model A for PAL was motivated by computational convenience, since this simpler case is sufficient to demonstrate the positive bias in the lookahead method for the 40-unit dataset.

A comparison of E_{14} and E_{15} shows the bias of PAL_L using the code and parameter values of WWR, changing only the number of particles. This provides the most direct evidence for H1, since the decrease of the log-likelihood by $\lambda_{14} - \lambda_{15} = -833.83$ log units when moving from $J_1 = 5 \times 10^3$ to $J_2 = 10^5$ suggests strongly that the log-likelihood estimate with $J_1 = 5 \times 10^3$ is over-optimistic.

Based on the comparison of E_4 and E_5, PAL_L does not report an accurate log-likelihood even on a single unit for $J = 10^5$, and 40 units is a considerably harder problem. Further, we know from E_8–E_{10} that PAL_L can over-state the true log-likelihood for the measles model with 40 units in a situation where the truth is known. The evidence suggests that neither E_{14} nor E_{15} is a reliable estimator of either the PAL log-likelihood or the exact, unknown, log-likelihood that PAL aims to approximate for the data. The performance of PAL_V on the 40-unit real dataset fails to outperform the ARMA benchmark, further indicating that applying PAL to high-dimensional data remains challenging.

3   Some theoretical considerations for lookahead PAL

Since vanilla particle filter algorithms are unbiased for the likelihood, it might be reasonable to expect the PAL-SMC algorithm to be unbiased for the PAL likelihood, but this is not true for the lookahead PAL filter used by WWR. This is a property of the lookahead part of the algorithm, developed by Rimella et al. (2023), rather than the PAL approximation. Therefore, for the remainder of this section, we consider the simpler lookahead filter of Rimella et al. (2023).

Briefly, the vanilla particle filter is unbiased for the likelihood (and, therefore, negatively biased for the log-likelihood) because the self-normalization denominator in the resampling step happens to coincide with the conditional likelihood estimate on the numerator, leading to a fortuitous cancellation. Self-normalization does not always lead to unbiased likelihood estimates, as we can see from the following example.

Let X take values $\{0, 1\}$ with equal probability, and let $Y = X$ with probability 1. Suppose a single data point, $Y = 1$. Suppose also an independent sample of J particles, $x_{1:J}$, each with distribution matching X. Now, resample these particle with probability $p_j = (1 - \epsilon)x_j/[\sum_j x_j] + \epsilon(1 - x_j)/[\sum_j 1 - x_j]$ so that, on average, a fraction $(1 - \epsilon)$ of the resampled particles have value 1. Take $\epsilon \ll 1/J$, so that most resampled particle swarms contain no particles with value 0. Most particle swarms resulting from resampling will have $x_j = 1$ for all j, with the proper resampled weight $w_j = 1/(J p_j)$ being approximately $1/(2J)$ for all j. Rare swarms will have a particle with massive weight, approximately $1/(2J\epsilon)$. Under self-normalization, particle

swarms with a massive weight will estimate the likelihood to be approximately zero, and particle swarms with $x_j = 1$ for all j will estimate the likelihood to be 1. By setting ϵ arbitrarily small, we can get an estimate whose expectation approaches 1 since with high probability we see only resampled swarms where every particle has value 1. If we take a different limit, with $J \to \infty$, the bias will go away asymptotically, but here we consider the case with fixed J and $\epsilon \to 0$.

Importantly, the bias on the likelihood estimate in this example is positive. As mentioned earlier, a suboptimal forecast generally gives, on average, a negative bias on the conditional log-likelihood estimate, since log-likelihood is a proper scoring rule. This justifies assessing filters on their log-likelihood estimate in a similar way that one does for parameters in likelihood-based inference. A filter with a high log-likelihood estimate on simulated data from the target model is validated as a good likelihood approximation. However, this does not necessarily apply to algorithms that look at future observations. When implementing lookahead algorithms, if you want the log-likelihood estimate to be conservative, you have to be extra careful to consider the bias. For unbiased likelihood estimates, the negative bias on the log-likelihood is a direct consequence of variance, and among such estimates it is reasonable to prefer a filter approximation with the highest log-likelihood estimate. For positively biased estimates, that is inappropriate.

4   Conclusion

The results in this article reinforce the investigation by Hao et al. (2024) and lead to the conclusion that there is not currently a strong case for using PAL on the epidemiological models used by WWR to motivate its development. Simpler particle filter methods apply to arbitrary Markov process models, whereas PAL is limited to a specific class of discrete-state Markov process models. Basic particle filters, and their block particle filter extensions, have the plug-and-play property (Bretó et al., 2009; He et al., 2010). Iterated block particle filters can be effective tools for data analysis of spatiotemporal epidemiological data (Wheeler et al., 2024; Li et al., 2024). Certain automatically differentiable particle filter algorithms inherit this convenient plug-and-play property while providing derivative estimates with low bias and variance (Tan et al., 2024), and these may in future be extended to block particle filters.

PAL may potentially lead to dramatic computational improvements over particle filters some situations. However, WWR's overdispersed generalization of PAL also involves a block particle filter component, at which point it shares many of the properties of particle filters. The results of WWR, together with various other authors (Stocks et al., 2018; He et al., 2010; Li et al., 2024), show that overdispersion is frequently necessary for a dynamic model to provide an adequate statistical description of epidemiological data.

References

Asfaw, K., Park, J., King, A. A., and Ionides, E. L. (2024). spatPomp: An R package for spatiotemporal partially observed Markov process models. *Journal of Open Source Software*, 9:7008.

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

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

Del Moral, P. (2004). *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Springer, New York.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378.

Hao, Y., Abkemeier, A. J., and Ionides, E. L. (2024). Poisson approximate likelihood compared to the particle filter. *arXiv:2409.12173*.

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

Ionides, E. L., Asfaw, K., Park, J., and King, A. A. (2023). Bagged filters for partially observed interacting systems. *Journal of the American Statistical Association*, 118(542):1078–1089.

Ionides, E. L., Ning, N., and Wheeler, J. (2024). An iterated block particle filter for inference on coupled dynamic systems with shared and unit-specific parameters. *Statistica Sinica*, 34:1145–1166.

Korevaar, H., Metcalf, C. J., and Grenfell, B. T. (2020). Structure, space and size: competing drivers of variation in urban and rural measles transmission. *Journal of the Royal Society Interface*, 17(168):20200010.

Li, J., Ionides, E. L., King, A. A., Pascual, M., and Ning, N. (2024). Inference on spatiotemporal dynamics for networks of biological populations. *Journal of the Royal Society Interface*, 21(216):20240217.

Ning, N. and Ionides, E. L. (2023). Iterated block particle filter for high-dimensional parameter learning: Beating the curse of dimensionality. *Journal of Machine Learning Research*, 24:1–76.

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

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

Rimella, L., Jewell, C., and Fearnhead, P. (2023). Approximating optimal SMC proposal distributions in individual-based epidemic models. *Statistica Sinica*, page SS–2022–0198.

Stocks, T., Britton, T., and Höhle, M. (2018). Model selection and parameter estimation for dynamic epidemic models via iterated filtering: Application to rotavirus in Germany. *Biostatistics*, 21(3):400–416.

Tan, K., Hooker, G., and Ionides, E. L. (2024). Accelerated inference for partially observed Markov processes using automatic differentiation. *arXiv:2407.03085*.

Wheeler, J., Rosengart, A., Zhuxun, J., Tan, K. H. E., Treutle, N., and Ionides, E. L. (2024). Informing policy via dynamic models: Cholera in Haiti. *PLOS Computational Biology*, 20:e1012032.

Whitehouse, M., Whiteley, N., and Rimella, L. (2023). Consistent and fast inference in compartmental models of epidemics using Poisson Approximate Likelihoods. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85(4):1173–1203.

Whitehouse, M., Whiteley, N., and Rimella, L. (2025). Correction to: Consistent and fast inference in compartmental models of epidemics using Poisson approximate likelihoods. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, pre-published online.

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