

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

Kunyang He, Yize Hao and Edward L. Ionides

2025-05-28

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 complex dynamic systems. 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 outperforms 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 positive bias in the log-likelihood estimates.

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 [REFS]. 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 dengue transmission model and a high-dimensional measles model. On close inspection, the dengue results turned out to be overstated [Hao et al., 2024] leading to a published correction [Whitehouse et al., 2025]. However, the measles results remained unchallenged, and our present goal is to revisit that evidence. In Section~?? 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~?? we explain theoretically how the positive bias for PAL arises as a result of the lookahead mechanism included in the implementation of PAL for this model. The lookahead mechanism was not used by WWR for the dengue analysis since its purpose was to address numerical issues involved in high-dimensional filtering. 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 simply provide an overview. The data are measles case counts aggregated over 2-week intervals for 40 of the largest towns in England and Wales, from 1949 to 1964. The data and the model are derived from Park and Ionides [2020] which builds on a long tradition of models described therein. Recently, weekly data for more towns have become publicly available [REF], 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 cities follows a power law with the distance between the towns, 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 comparing filters on spatiotemporal measles models. Here, we focus on developing experiments aimed at establishing two specific hypotheses:

- Q1 The lookahead version of the Poisson approximate likelihood estimator of WWR (PAL_L) can have substantial positive bias on its log-likelihood estimate.
- Q2 The bias scales 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 which 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, these being UK towns in the measles example. The experimental variables, and the list of values we consider for them, are summarized in Table~?? and further described below.

A critical 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 with a small known bound on the error. That is the reason why algorithms such as (PAL_L) are being invented. There are two special situations where we can establish 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 it is enough to solve the filtering problem exactly for a 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 [?] 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, 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.

Evidently, the basic particle filter is not a powerful tool for general spatiotemporal systems, and so alternative approaches are needed. The filtering and prediction problems are useful for situations where we do not have to make a forecast, and so the filter’s prediction distribution for the n th observation, at time t_n , can be designed taking advantage of data occurring at, or subsequent to, time t_n . This can allow the construction of numerically efficient algorithms, and that is the goal of the lookahead filter, PAL_L . 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

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	$\hat{\theta}_{BPF}^*$	$\hat{\theta}_{PAL}$	

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

to construct the filter. We will investigate the theory behind $\{\text{PAL}_L\}$ later, in Section~???, but for now we just need to note the potential hazard.

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 block particle filter algorithm [[Ning and Ionides, 2023, ?](#)] 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 `spatPomp::bpfilter` even when `pomp::pfilter` is equivalent.

f_C Spatiotemporal coupling for the filter model. 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 ($\iota \neq 0$). Note that when $U = U_1 = 1$, we use the largest city, London, for which stochastic extinctions are very unlikely. Also, for $U = 1$, the value of g becomes irrelevant.

f_{proc} Latent process transition model for the filter. $f_{\text{proc}} = \text{discrete}$ corresponds to the choice of WWR where a single gamma-distributed dynamic noise variable is chosen for each observation interval. $f_{\text{proc}} = \text{Euler}$ corresponds to independent gamma noise variables on 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 $f_{\text{proc}} = \text{discrete}$ and $f_{\text{proc}} = \text{Euler}$ 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. $f_{\text{meas}} = \text{binomial}$ corresponds 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. $f_{\text{meas}} = \text{Gaussian}$ corresponds to a discretized Gaussian measurement model. Using a measurement model that can directly be evaluated, without costly Monte Carlo calculation, assists with efficient Monte Carlo inference. Beta-binomial or negative binomial distributions would also meet this criterion.

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

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_L	J_1	U_1	C_1	discrete	binomial	$\hat{\theta}_{PAL}$	—————	data	—————	—————

Table 2: Combinations of variable values used for each experiment, E_k , $k = 1 \dots 13$. The simulation settings, s_C , s_{proc} , s_{meas} and s_θ , are applicable only when we filter simulated data rather than real data. *NOTE: THIS TABLE IS PRELIMINARY. IT NEEDS CAREFUL PROOF READING. POSSIBLE COPY-PASTE MISTAKES!*

provided by WWR.

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

s_{meas} Measurement model for the simulation. Always set to $s_{\text{meas}} = \text{Gaussian}$ 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 exact value of s_θ is not critical to the argument, but we use a value obtained by maximizing the BPF likelihood.

We consider two software platforms for the experiments: all $\{\text{PAL}_L\}$ and $\{\text{PAL}_V\}$ calculations are carried out using the code provided by WWR, and all BPF calculations are carried out using the pomp [King et al., 2016] and spatPomp R packages.

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~TODO, together with log-ARMA and autoregressive negative binomial benchmarks.

Here, experiments E_1 – E_5 address Q1 directly, using simulated data on a single unit. E_1 provides a ground truth for a particular model. E_4 and E_5 demonstrate positive bias from $\{\text{PAL}_L\}$ both at a usual number of particles and for an intensive calculation (which may not be possible on larger problems). This calculation imposes a negative bias on both PAL algorithms because the true model differs from the model underlying PAL. E_2 and E_3 are non-forward-looking algorithms which provide an estimate of the PAL likelihood; E_4 and E_5 are theoretically consistently estimating the same quantity as E_2 and E_3 but hopefully with reduced variance. Thus, the comparison between E_2 and E_4 is our best estimate of the positive bias of $\{\text{PAL}_V\}$, when conducted with a moderate number of particles, in this case.

Experiments E_6 and E_7 introduce the actual data. Comparing the resulting log-likelihoods with the benchmark values, we see **TO DO. I EXPECT TO SEE TYPICAL VALUES FOR HOW THE METHODS COMPARE TO THE BENCHMARKS ON REAL AND SIMULATED DATA, FOR ONE UNIT**

Experiments E_8 provides a 40 unit system for which a ground truth is known. E_{10} shows that $\{\text{PAL}_L\}$ may over-estimate the truth, whereas E_9 shows that $\{\text{PAL}_V\}$ does not.

Experiments E_{11} – E_{13} consider coupled models for the full, real dataset. We cannot accurately know the ground truth here. However, comparing estimated likelihood with the benchmarks, in light of the earlier experiment, the results are consistent with a conclusion that... TODO

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, derived from [REF], 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 because the self-normalization constant happens to coincide with the conditional likelihood estimate. 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 $J \rightarrow \infty$, the bias will go away asymptotically, but here we consider the case with fixed J and $\epsilon \rightarrow 0$.

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 results by Hao et al. [2024] and lead to the conclusion that there is not currently a strong case for using PAL. PAL applies to a specific class of discrete-state Markov process models, whereas simpler particle filter methods apply much more generally. Basic particle filters, and their block particle filter extensions, have the plug-and-play property [Bretó et al., 2009, He et al., 2010] and likelihood optimization methods for particle filters such as iterated filtering [Ionides et al., 2015] and certain automatic differentiation algorithms [Tan et al., 2024] inherit this convenient property. PAL may potentially lead to dramatic computational improvements over particle filters some situations. However, WWR’s overdispersed

generalization of PAL also requires a particle filter component, at which point it shares the limitations of particle filters. WWR, together with various other authors [Stocks et al., 2018, He et al., 2010], found that overdispersion is necessary for the model to provide an adequate statistical description of epidemiological data.

References

- K. Asfaw, J. Park, A. A. King, and E. L. Ionides. spatPomp: An R package for spatiotemporal partially observed Markov process models. *Journal of Open Source Software*, 9:7008, 2024.
- C. Bretó, D. He, E. L. Ionides, and A. A. King. Time series analysis via mechanistic models. *Annals of Applied Statistics*, 3:319–348, 2009.
- P. Del Moral. *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Springer, New York, 2004.
- Tilman Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.
- Yize Hao, A. J. Abkemeier, and Edward L. Ionides. Poisson approximate likelihood compared to the particle filter. *arXiv:2409.12173*, 2024.
- D. He, E. L. Ionides, and A. A. King. 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, 2010.
- Edward L. Ionides, Dao Nguyen, Yves Atchadé, Stilian Stoev, and Aaron A. King. Inference for dynamic and latent variable models via iterated, perturbed Bayes maps. *Proceedings of the National Academy of Sciences of USA*, 112(3):719–724, 2015. doi: 10.1073/pnas.1410597112.
- Edward L. Ionides, Kidus Asfaw, Joonha Park, and Aaron A. King. Bagged filters for partially observed interacting systems. *Journal of the American Statistical Association*, 118(542):1078–1089, 2023. doi: 10.1080/01621459.2021.1974867. URL <https://doi.org/10.1080/01621459.2021.1974867>.
- Edward L. Ionides, Ning Ning, and Jesse Wheeler. An iterated block particle filter for inference on coupled dynamic systems with shared and unit-specific parameters. *Statistica Sinica*, 34:1145–1166, 2024.
- Aaron A. King, Dao Nguyen, and Edward L. Ionides. Statistical inference for partially observed Markov processes via the R package pomp. *Journal of Statistical Software*, 69(12):1–43, 2016.
- Ning Ning and Edward L Ionides. Iterated block particle filter for high-dimensional parameter learning: Beating the curse of dimensionality. *Journal of Machine Learning Research*, 24:1–76, 2023.
- J. Park and E. L. Ionides. Inference on high-dimensional implicit dynamic models using a guided intermediate resampling filter. *Statistics & Computing*, 30:1497–1522, 2020.
- Patrick Rebeschini and Ramon Van Handel. Can local particle filters beat the curse of dimensionality? *The Annals of Applied Probability*, 25(5):2809–2866, 2015.
- Lorenzo Rimella, Christopher Jewell, and Paul Fearnhead. Approximating optimal SMC proposal distributions in individual-based epidemic models. *Statistica Sinica*, page SS–2022–0198, 2023.
- Theresa Stocks, Tom Britton, and Michael Höhle. Model selection and parameter estimation for dynamic epidemic models via iterated filtering: application to rotavirus in Germany. *Biostatistics*, 21(3):400–416, 09 2018.
- Kevin Tan, Giles Hooker, and Edward L. Ionides. Accelerated inference for partially observed Markov processes using automatic differentiation. *arXiv:2407.03085*, 2024.

Michael Whitehouse, Nick Whiteley, and Lorenzo Rimella. 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, 07 2023.

Michael Whitehouse, Nick Whiteley, and Lorenzo Rimella. 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, 2025.