

The Poisson approximate likelihood compared to the particle filter

Yize Hao, Aaron A. Abkemeier and Edward L. Ionides

Draft compiled on August 18, 2024

Abstract

Filtering algorithms are fundamental for inference on partially observed stochastic dynamic systems, since they provide access to the likelihood function and hence enable likelihood-based or Bayesian inference. 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 and consistent approximation to the likelihood function for a certain subset of partially observed Markov process models. [Whitehouse et al. \(2023\)](#) supported the importance of their contribution by theoretical analysis and numerical results. A central piece of evidence for PAL is the comparison in Table 1 of [Whitehouse et al. \(2023\)](#), which claims a large improvement for PAL over a standard particle filter algorithm. This evidence, based on a model and data from a previous scientific study by [Stocks et al. \(2018\)](#), might suggest that researchers confronted with similar models should use PAL rather than previous particle filter methods. Alternatively, the improvement in likelihood found by [Whitehouse et al. \(2023\)](#) compared to [Stocks et al. \(2018\)](#) might indicate a weakness in that previous work. We show that the comparison of log-likelihood values made by [Whitehouse et al. \(2023\)](#) is flawed because the PAL calculations were carried out using a dataset scaled differently from the previous study. If PAL and the particle filter are used on this new scale, the superficial advantage of PAL largely disappears. On simulations where the model is correctly specified, the particle filter outperforms PAL. If a poorly fitting specification of initial value parameters is amended, the particle filter also outperforms PAL on the actual data.

1 Introduction

This article results from an investigation of the results in Table 1 of [Whitehouse et al. \(2023\)](#). The authors of that paper were given the opportunity to submit a correction, after we shared the results of our investigation with them, but they declined. The theory developed by [Whitehouse et al. \(2023\)](#) shows that PAL has some potentially useful scaling properties, but the numerical results in Table 1 appear to show much stronger performance than a standard particle filter on this example

of moderate size. Our task here is to correct the error in Table 1 so that researchers considering whether to implement Poisson approximate likelihood (PAL) are appropriately informed about its benefits. Table 1 reanalyzes the model and data of [Stocks et al. \(2018\)](#), for which the likelihood was calculated using a particle filter. [Stocks et al. \(2018\)](#) found strong evidence for the importance of overdispersion in a stochastic dynamic model for their epidemiological data. This is significant because earlier research on population dynamics largely avoided consideration of overdispersion, perhaps because of the lack of statistical methodology to fit such models. The conclusions of [Stocks et al. \(2018\)](#) hinge on a comparison of likelihoods, and so the results of [Whitehouse et al. \(2023\)](#) discredit those conclusions by indicating that [Stocks et al. \(2018\)](#) used an inaccurate filter to calculate their likelihoods. An important consequence of correcting Table 1 is that the results of [Stocks et al. \(2018\)](#) stand undiminished.

First, we show that most of the apparent advantage for PAL, compared to a particle filter, arises because [Whitehouse et al. \(2023\)](#) used a different scaling of the data from [Stocks et al. \(2018\)](#). Two models for the same data can properly be compared by their likelihood, even if the models have entirely different structures. One can make allowance for the number of estimated parameters using a quantity such as Akaike’s information criterion ([Akaike, 1974](#)). However, if data are rescaled, a correction is required to make likelihoods comparable. For example, if one model describes a dataset in grams and another describes it in kilograms, then the latter model will earn an increased log-likelihood of $\log(10^3)$ for each data point simply because of the change in scale. Presenting a direct comparison of a likelihood for the data in grams with a likelihood for the data in kilograms would evidently be inappropriate. [Stocks et al. \(2018\)](#) fitted their model to a dataset derived by dividing the original reported count data by an estimated reporting rate, to put their data on the scale of the actual number of cases in the population, whereas [Whitehouse et al. \(2023\)](#) fitted directly to the report data. The reporting rate used by [Stocks et al. \(2018\)](#) varied over time, but was generally around 7%. On approximately 1200 data points, this corresponds to a discrepancy of around $-1200 \log(0.07) \approx 3200$ log-likelihood units, largely explaining the difference reported in Table 1 and interpreted by [Whitehouse et al. \(2023\)](#) as evidence supporting PAL. The comparison can be corrected either by applying the method of [Stocks et al. \(2018\)](#) to the data of [Whitehouse et al. \(2023\)](#) or vice versa. Since the method of [Stocks et al. \(2018\)](#) is applicable to a more general class of models, and supported by published software, it was more convenient to apply this method to the model and data of [Whitehouse et al. \(2023\)](#). The large discrepancy in log-likelihood disappears at this point by recomputing the likelihood of the model using PAL and particle filter separately (see Table 1). This re-analysis does, however, show a discrepancy between two methods. We continued our investigation to establish the cause of this.

PAL provides an approximate filter, whereas the particle filter provides a Monte Carlo estimate of an exact filter. Possible explanations for PAL obtaining a higher log-likelihood than a particle filter include:

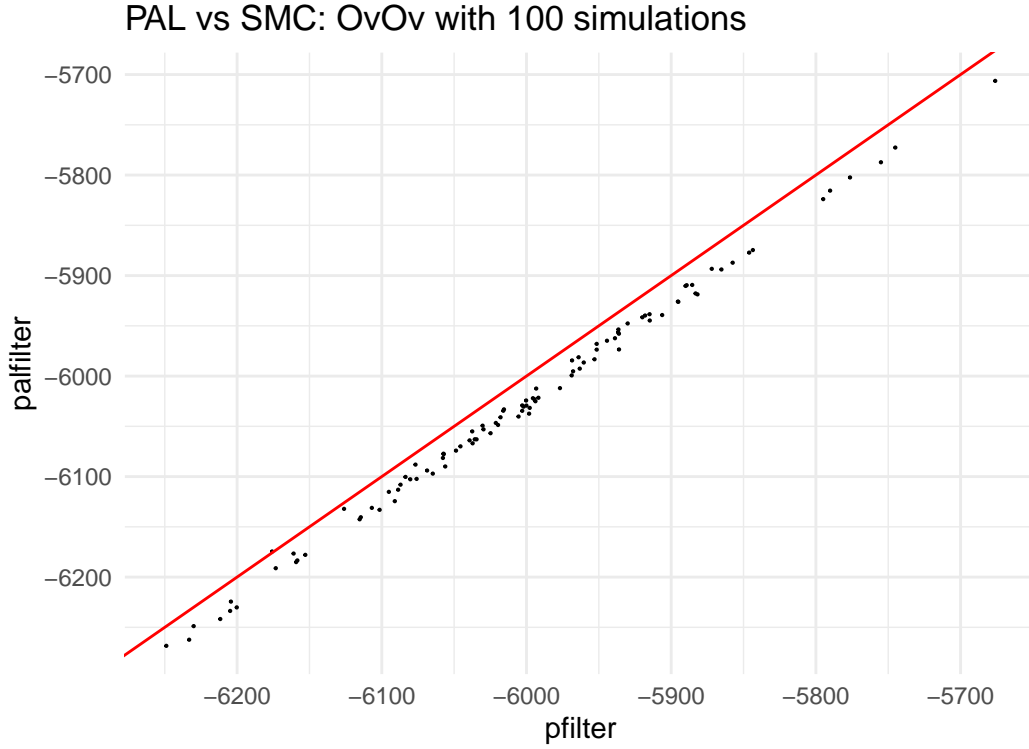


Figure 1: Log-likelihoods computed using two filtering methods for 100 randomly simulated datasets at the MLE of the OvOv model. The particle filter and PAL used 50,000 particles. The red line corresponds to equality of the two estimates. On average, the particle filter likelihood estimate is 7.7 log units higher than the PAL.

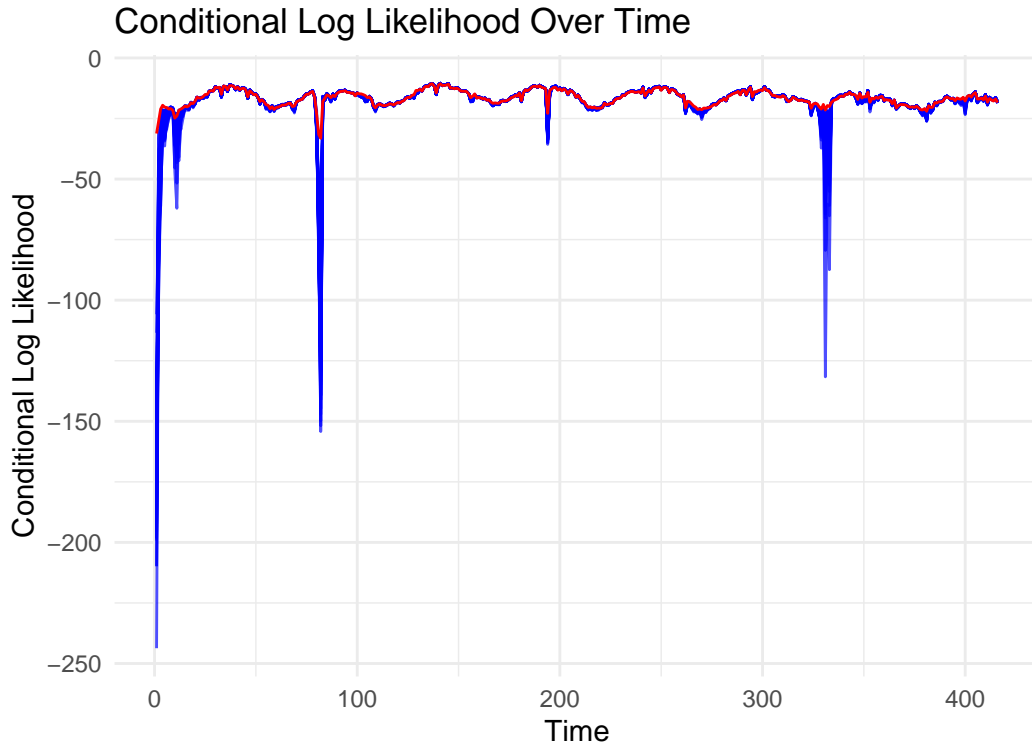


Figure 2: Conditional log-likelihoods computed using two methods for the real rotavirus dataset, with the SMC method replicated 36 times due to its high variances. The blue line represents overlapping results from 36 SMC computations, while the red line is derived from PAL. The main sources of likelihood shortfall of SMC are at time points $t = 1, 2, 3, 11, 81, 194$, and $325 - 333$.

1. The particle filter could have high Monte Carlo variance. This results in negative bias in its log-likelihood estimate due to Jensen’s inequality, since the particle filter provides an unbiased likelihood estimate.
2. If the model is misspecified, the approximation error in PAL could lead to a higher log-likelihood estimate than that of the actual misspecified model.

Log-likelihood is a proper scoring rule for forecasts [Gneiting and Raftery \(2007\)](#), and both the particle filter and PAL construct their log-likelihood estimates via a sequence of one-step forecasts. Therefore, if the model is correctly specified, the approximation error in PAL can only decrease the expected log-likelihood. We tested this on simulated data, for which the model of [Whitehouse et al. \(2023\)](#) is correctly specified. We focus on the version of the model with overdispersion in both the population dynamics and the measurement process. This version of the model, called OvOv, was found by both [Whitehouse et al. \(2023\)](#) and [Stocks et al. \(2018\)](#) to be the best choice to fit the data. For this simulation study, the particle filter out-performs PAL (Figure 1) which supports hypothesis 2, above. Consequently, we looked for model misspecification by comparing the conditional log-likelihood estimates at each time point, as described in the Methods section.

Table 1: log-likelihood at the MLE of OvOv model for the real rotavirus data, computed using two filtering methods, employed 36 replicates and 50,000 particles

	PAL	SMC
OvOv	-6892.3	-7268

Figure 2 indicates that the worst conditional log-likelihood values, estimated using the particle filter, arise at the start of the time series. This could be a problem with data collection, or some subtle issue with the science of the epidemiological system, but a simple possibility is that the initial values of the latent state process might be poorly specified. For this model, the initial values were fixed based on scientific reasoning, rather than being estimated. If we instead estimate the initial values, we obtain the results in Table 2.

[EI:WE CAN REVIEW EXACTLY WHICH NUMBERS TO SHOW. IDEALLY, AS LITTLE AS POSSIBLE IN ORDER TO MAKE A STRONG ARGUMENT.]

We conclude that PAL is a potentially useful algorithm, with some favorable theoretical properties. However, the corrected evidence does not indicate an advantage for using PAL in situations where the particle filter is effective.

Methods

[EI:BRIEFLY DESCRIBE HERE ANY DETAILS NEEDED TO DESCRIBE THE RESULTS. COMMENT ON CODE AVAILABILITY.]

[YH:ON AVERAGE PAL IS 25.3 LOG UNITS LOWER THAN PFILTER AND SHOULD WE

Table 2: Maximum likelihood estimation of parameters by PAL ([Whitehouse et al., 2023](#)), and by mif2 algorithm ([King et al., 2016](#)) for the OvOv model.

Parameter	OvOv	OvOv (pomp)*
β_1	11.48	9.84
β_2	0.25	0.24
β_3	0.35	0.74
ϕ	0.14	0.17
ρ	0.16	0.18
σ_q^2	0.021	0.019
σ_ξ	66.89	13.96
S_{10}	3876549	6145778
I_{10}	30351	2181
R_{10}	1315221	468260
S_{20}	57139612	37469849
I_{20}	871	1942
R_{20}	302852	31572281
S_{30}	19573727	4540574
I_{30}	2550	774
R_{30}	131092	2171187
AIC	13778.08	13630.26

* The initial distribution parameter $\lambda_0 = (S_{10}, I_{10}, R_{10}, S_{20}, I_{20}, R_{20}, S_{30}, I_{30}, R_{30})$ are assumed to be **fixed** in [Whitehouse et al. \(2023\)](#). The results of the estimates of the OvOv (**pomp**) model are obtained by using the Iterated Filtering algorithm with 3 rounds and 100 iterations, 50,000 particles, and 36 replicates in each round with the top 12 best fits in terms of likelihood chosen to be the starting value for the next round.

MENTION THAT WE USED THE SIMULATED DATA WITHOUT ZEROS?

THE TABLE SEEMS TO BE TALL AND THIN. SHOULD WE PUT IT IN A APPENDIX?]

To evaluate whether SMC performs as well as PAL in terms of both computational time and likelihood identifiability, we replicated the model using the R package **pomp** (King et al., 2016). First, we verified that the two models are essentially identical. Here, we simulated data from the two models 1000 times at the MLE shown in Table 2 and compared the summary statistics of 1000 simulated time series from two models using the t-test. Summary statistics including the mean, median, and variance for all three age groups were tested. The results showed no significant differences between the two models from which the data were simulated, confirming that both the latent process model and the observation model are essentially consistent with each other, given that the observations were directly simulated. From Table 2, it is evident that the OvOv model, with overdispersion in both the dynamic system and the measurement model, yields the lowest AIC. We therefore focus on this version of the model.

This article has focused on likelihood evaluation, but the inferences presented also require the likelihood estimate to be maximized. Stocks et al. (2018) used iterated filtering to maximize the particle filter likelihood estimate, whereas Whitehouse et al. (2023) used direct numerical maximization. We follow the approaches adopted by these two papers. The optimization method is not directly pertinent to the comparison of the filtering methods, so we do not discuss optimization in further detail here.

Our workflow is fully reproducible. An extended description of our reanalysis of Whitehouse et al. (2023) is provided by Hao (2024). That thesis also contains some additional results reinforcing the conclusions presented in this article.

References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716–723.
- 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. (2024). PAL versus SMC: Two approaches in compartmental modeling. a case study with rotavirus in Germany during 2001-2008. University of Michigan Honors Thesis.
- King, A. A., Nguyen, D., and Ionides, E. L. (2016). Statistical inference for partially observed Markov processes via the R package pomp. *Journal of Statistical Software*, 69(12):1–43.
- 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.

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.