PAL versus SMC: Two Approaches in Compartmental Modeling

A case study with rotavirus in Germany during 2001-2008

Yize Hao

Supervisor: Prof. Edward Ionides

Graduate Student Mentor: Aaron Abkemeier

Submitted: April 23, 2024

This version: May 31, 2024

Abstract

Partially Observed Markov Process (POMP) models have been extensively employed in epidemiological modeling over the past several decades to understand disease patterns and inform policy-making. Although the observation density is typically assumed to be known or at least can be evaluated point-wisely, the latent Markovian models lack closed-form expressions for initial density and transition kernels, making it challenging to compute the likelihood for POMP models with complex latent models. Recently, a novel approach (Poisson Approximate Likelihood, PAL) was introduced by Whitehouse et al. (2023), which employs a Poisson approximation to posterior densities, offering a fast and consistent approximation for the likelihood function. Whitehouse et al. (2023) claimed that their method, along with its associated model, improved the maximum likelihood estimation compared to traditional sequential Monte Carlo (SMC) methods used in Stocks et al. (2018) when applied to the German rotavirus, by approximately 3200 log-likelihood units. However, our analysis of comparing two methods applied to the model for the same rotavirus dataset reveals that the improvement of 3200 log-likelihood units results from the use of two different datasets differed by a scaling factor. Moreover, although PAL and SMC are two ways of approximating the likelihood, within the framework of models compatible with PAL, when computation time is preferred, PAL is recommended but it may suffer from a likelihood shortfall, which can't be overcome in general. When the model is misspecified for certain time points, SMC may fail, and PAL is possible to approximate the likelihood. When computational time is not a critical factor and the model is correctly specified, sequential Monte Carlo methods are recommended.

Keywords: epidemiology; compartmental model; state-space model; sequential Monte Carlo; likelihood-based inference

1 Introduction

Understanding dynamic diseases is critical for public health and policy-making. Traditional time series models, such as the TSIR model (Bjørnstad et al., 2002; Becker and Grenfell, 2017), are valuable for predicting and forecasting the transmission scale of specific infectious diseases, as well as for estimating critical parameters such as the basic reproduction number R_0 and the force of infection. However, these models still fail to provide a sufficiently clear and straightforward understanding of the disease's transmission dynamics. Therefore, modeling the epidemiology data differently is necessary. Here, we seek an alternative approach — Partially Observed Markov Process (POMP) models with compartmental modeling in the latent process, which is one of the most widespread models for understanding the dynamics of infectious diseases and epidemiology data in the recent several decades (Michael A. Newton and Abkowitz, 1995; Hulin and Wai-Yuan, 2000; He et al., 2010; Doan et al., 2015; Stocks et al., 2018; Jiang et al., 2021; He et al., 2023; Whitehouse et al., 2023). However, when there is a desire for models to more closely resemble reality—namely, moving beyond linear Gaussian assumptions—traditional filtering methods such as the Kalman Filter (Kalman, 1960) and the Ensemble Kalman Filter (Evensen, 1994; Houtekamer and Mitchell, 1998) become unsuitable. This complexity arises because many diseases are inherently more complicated; they are often modeled as nonlinear, non-Gaussian, high-dimensional, and highly stochastic POMP models with several compartments in the latent process.

POMP models are defined by their initial conditions, transition kernels, and observation densities. Though the observation density is always assumed to be mathematically evaluatable, the latent Markovian models have neither a closed-form initial density nor transition kernels. Therefore, the likelihood often is hard to retrieve when such a model is fairly complicated. However, as proposed by Kitagawa (1987); Gordon et al. (1993); Kitagawa (1996), Monte Carlo methods can be applied sequentially to approximate the posterior distribution, as well as the likelihood function. This is the so-called and widely-used sequential Monte Carlo method for filtering general models as long as one can simulate from the latent Markovian model, which is based on the idea of Sampling-Importance-Resampling (Rubin, 1987).

Owing to the unique characteristics of POMP models, which we will introduce subsequently, these models have gained significant importance and have been extensively applied in engineering and related disciplines, such as speech word identification (Rabiner, 1989), target positioning and velocity estimation in radar systems (Gordon et al., 1993; Avitzour, 1995), volatility estimation in economic time series (Pitt and Shephard, 1999), and structural type prediction in protein secondary structures (Schmidler et al., 2000). In these models, the unobserved signal process $\{x_n : n \in \mathbb{N}\}, x_n \in \mathcal{X}$, represents the true underlying signal, where $\mathcal{X}$ denotes the space where the state vectors live. A key example is the widely utilized SIR model for infectious disease modeling, which categorizes individuals into three compartments: S (susceptible), I (infectious), and R (recovered) (He et al., 2010; Battineni et al., 2020; Biswas et al., 2014). In this case, $\mathcal{X} = \mathbb{N}^3$, where $\mathbb{N}$ is the set of natural numbers. $x_n = (S_n, I_n, R_n)$ where S_n, I_n, and R_n indicate the number of individuals in each compartment at time n. However, what is ob-

servable is $\{y_n | n \in \mathbb{N}\}, y_n \in \mathcal{Y}$, which depends only on $\{x_n : n \in \mathbb{N}\}, x_n \in \mathcal{X}$ and always includes noise components. In the above SIR example, $\mathcal{Y} = \mathbb{N}$ since y_n, our data, is the observed number of cases at time n. Therefore, accurately retrieving the values of $\{x_n : n \in \mathbb{N}\}, x_n \in \mathcal{X}$ based on the observations $y_{1:N}$ is crucial, a challenge commonly referred to as a filtering problem. This entails filtering out the noise to recover the posterior distribution of $x_{1:n}$ given $y_{1:n}, n \in \mathbb{N}$, a problem widely considered in Bayesian statistics. (Gordon et al., 1993; Geweke, 1989; Doucet et al., 2000; Andrieu et al., 2010)

When inference and estimation of parameters are of interest, likelihood maximization methods are particularly challenging when the likelihood function is mathematically intractable, as detailed in Bretó et al. (2019); Ionides et al. (2011); He et al. (2010). These methods often rely solely on the "plug-and-play" property, which depends only on simulations from the latent process model. Conversely, when the likelihood function is accessible, either through integration or approximation, methods such as the gradient ascent algorithm (Lemaréchal, 2012; Hadamard, 1908; Courant, 1943) and the Stochastic EM algorithm (Nielsen, 2000) can be effectively utilized.

Next, we introduce the POMP model in general and then review two different filtering methods that were proposed and compared in Whitehouse et al. (2023). Subsequently, we will discuss the model specifically used for the German rotavirus from 2001 to 2008 as a case study, which will enhance understanding of the two methods and, more importantly, provide guidance on whether and when one should use them.

2 Methodology

Notation	Description
θ, ϕ	Parameters, usually $\theta, \phi \in \mathbb{R}^p, p \geq 1$
$f(\cdot; \theta), p(\cdot; \theta)$	Probability density functions parametrized by parameter θ of continuous and discrete random variables respectively
$X_{1:n}$	State variable from time 1 to n.
$Y_{1:n}$	Observation variable from time 1 to n.
$x_{1:n}$	The realization of the state variable from time 1 to n.
$y_{1:n}$	The realization of the observation variable from time 1 to n.
$\mathcal{L}(\theta|\cdot) = f(\cdot|\theta)$	The likelihood function at parameter θ given data.
$\ell(\theta|\cdot) = log f(\cdot|\theta)$	The log-likelihood function at parameter θ given data.

2.1 The Partially Observed Markov Process (POMP) Model

The Partially Observed Markov Process (POMP) model (He et al., 2010), also known as the state-space model (SSM), or the Hidden Markov Model (HMM), is a time series model consisting of processes (Chopin and Papaspiliopoulos, 2020; Liu, 2008; Doucet et al., 2001): A latent process/ state model, written as $x_n \sim f_{X_n|X_{n-1}}(\cdot|x_{n-1}, \theta)$ and an observation/ measurement model, written as $y_n \sim f_{Y_n|X_n}(\cdot|x_n, \theta)$. Here,

we only consider the discrete-time representation. Where $\{x_n : n \in \mathbb{N}\}, x_n \in \mathcal{X}$ denotes the hidden or latent states representing the unobserved signal. These states are treated as missing data within our model and are of interest, a point that will be elaborated upon subsequently. And $\{y_n | n \in \mathbb{N}\}, y_n \in \mathcal{Y}$ are called observations and are our actual observed data. The latent process is modeled as a Markov process, i.e. $\{x_n : n \in \mathbb{N}\}$ has the Markovian property: $f_{X_n|X_{1:n-1}}(x_n|x_{1:n-1}, \theta) = f_{X_n|X_{n-1}}(x_n|x_{n-1}, \theta)$ with initial distribution $f_{X_0}(x_0; \theta)$ and the transition kernel $f_{X_n|X_{n-1}}(x_n|x_{n-1}, \theta)$. $\{y_n | n \in \mathbb{N}\}$ are conditionally independent given the state variables $\{x_n : n \in \mathbb{N}\}$, and if $\{x_n : n \in \mathbb{N}\}$ were given, y_n only depends on x_n in the way that $y_n \sim f_{Y_n|X_n}(y_n|x_n, \theta)$. And $f_{Y_n|X_n}(y_n|x_n, \theta)$ is called the observation/ measurement density. Such conditions are visualized by Figure 1.

Consequently, the likelihood function can be written as:

$$\mathcal{L}(\theta|y_{1:N}) = f_{Y_{1:N}}(y_{1:N}; \theta) \tag{1}$$

$$= \int f_{Y_{1:N}, X_{0:N}}(y_{1:N}, x_{0:N}; \theta) dx_{0:N} \tag{2}$$

$$= \int f_{X_0}(x_0; \theta) \prod_{n=1}^{N} f_{Y_n|X_n}(y_n|x_n, \theta) f_{X_n|X_{n-1}}(x_n|x_{n-1}, \theta) dx_{0:N} \tag{3}$$
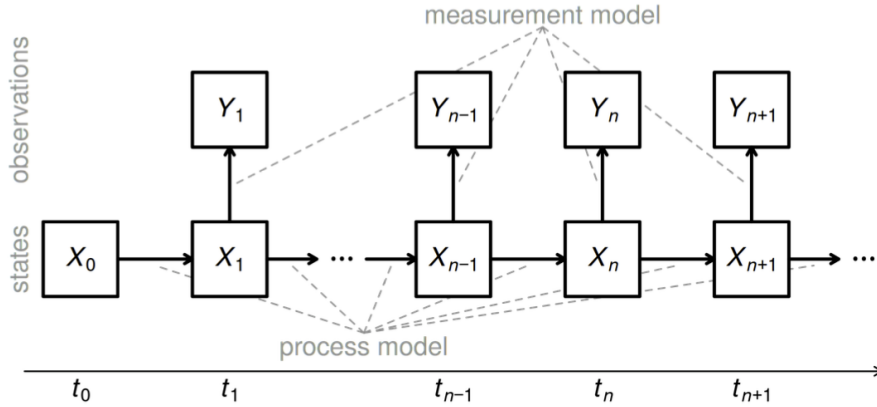

Figure 1: Partially Observed Markov Process (POMP) Model (King, 2024)

However, $f_{X_0}(x_0; \theta)$, and $f_{X_n|X_{n-1}}(x_n|x_{n-1}; \theta)$ doesn't always have a closed-form expression, especially for compartmental modeling for complex epidemiology data. Therefore, evaluating the likelihood by simply calculating the integral directly is nearly impossible. Here, for estimation and inference, we review two methods for calculating the likelihood.

2.2 Sequential Monte Carlo/ Particle Filtering (SMC)

The sequential Monte Carlo (SMC), also known as Particle Filtering, has been widely used with various variants developed over recent decades for posterior density approximation, including $f_{X_n|Y_{1:n-1}}(x_n|y_{1:n-1}, \theta)$ and $f_{X_n|Y_{1:n}}(x_n|y_{1:n}, \theta)$, as well as for likelihood approximation. Theoretical foundations are discussed in Chopin and Papaspiliopoulos (2020); Doucet et al. (2001); Liu (2008), and R packages implementing SMC are detailed in King et al. (2016). Here, we introduce the basic concept of how SMC functions for

likelihood approximation to serve our specific objectives.

Suppose we are interested in computing the likelihood for such POMP models. Recall that:

$$\mathcal{L}(\theta|y_{1:N}) = f_{Y_{1:N}}(y_{1:N};\theta)$$

$$= \prod_{n=1}^{N} f_{Y_n|Y_{1:n-1}}(y_n|y_{1:n-1},\theta)$$

$$= \prod_{n=1}^{N} \mathcal{L}_n(\theta)$$

where we let $f_{Y_n|Y_{1:n-1}}(y_n|y_{1:n-1},\theta) = f_{Y_1}(y_1|\theta)$, when $n = 1$, and $\mathcal{L}_n(\theta) := f_{Y_n|Y_{1:n-1}}(y_n|y_{1:n-1},\theta)$. Thus, we merely need to compute $\mathcal{L}_n(\theta)$. By the following formula:

$$\mathcal{L}_n(\theta) = f_{Y_n|Y_{1:n-1}}(y_n|y_{1:n-1},\theta)$$

$$= \int f_{Y_n,X_n|Y_{1:n-1}}(y_n,x_n|y_{1:n-1},\theta)dx_n$$

$$= \int f_{Y_n|X_n,Y_{1:n-1}}(y_n|x_n,y_{1:n-1},\theta)f_{X_n|Y_{1:n-1}}(x_n|y_{1:n-1},\theta)dx_n$$

$$= \int f_{Y_n|X_n}(y_n|x_n,\theta)f_{X_n|Y_{1:n-1}}(x_n|y_{1:n-1},\theta)dx_n$$

$$= \mathbb{E}_{X_n|Y_{1:n-1}}[f_{Y_n|X_n}(y_n|X_n,\theta)|y_{1:n-1},\theta]$$

by the conditional independency of y_n given x_n. Then if we can sample $x_n^{(i)} \sim p(x_n|y_{1:n-1},\theta)$, by Law of Large Numbers, we have

$$\frac{1}{J}\sum_{i=1}^{J} f_{Y_n|X_n}(y_n|x_n^{(i)},\theta) \xrightarrow{J\to\infty} \int f_{Y_n|X_n}(y_n|x_n,\theta)f_{X_n|Y_{1:n-1}}(x_n|y_{1:n-1},\theta)dx_n$$

$$= f_{Y_n|Y_{1:n-1}}(y_n|y_{1:n-1},\theta)$$

Therefore, our objective shifts to sequentially sampling $x_n^{(i)}$ from $f_{X_n|Y_{1:n-1}}(x_n|y_{1:n-1},\theta)$ for each $n \geq 1$. Sampling sequentially means that we don't want to sample a whole different set of $\{x_{n+1}^{(i)}\}_i$ whenever a new observation y_{n+1} is available. This can be achieved by the sequential Monte Carlo technique as shown in Algorithm 1.

Algorithm 1 Sequential Monte Carlo (SMC, or particle filter)

Input: Simulator for $f_{X_n|X_{n-1}}(x_n|x_{n-1};\theta)$; evaluator for $f_{Y_n|X_n}(y_n|x_n;\theta)$; simulator for $f_{X_0}(x_0;\theta)$; parameter, θ; data, $y_{1:N}$; number of particles, J.

1: **for** n in $1:N$ **do**
2:     Simulate for prediction: $X_{n,j}^P \sim f_{X_n|X_{n-1}}(\cdot|X_{n-1,j}^F;\theta)$ for j in $1:J$.
3:     Evaluate weights: $w(n,j) = f_{Y_n|X_n}(y_n^*|X_{n,j}^P;\theta)$ for j in $1:J$.
4:     Normalize weights: $\tilde{w}(n,j) = \frac{w(n,j)}{\sum_{m=1}^{J} w(n,m)}$.
5:     Apply select indices $k_{1:J}$ with $P[k_j = m] = \tilde{w}(n,m)$.
6:     Resample with replacement: set $X_{n,j}^F = X_{n,k_j}^P$ for j in $1:J$.
7:     Compute conditional log likelihood: $\hat{l}_{n|1:n-1} = \log\left(J^{-1}\sum_{m=1}^{J} w(n,m)\right)$.
8: **end for**
output: Log likelihood estimate, $\hat{l}(\theta) = \sum_{n=1}^{N} \hat{l}_{n|1:n-1}$; filter sample, $X_{n,1:J}^F$, for n in $1:N$.
complexity: $O(J)$

Moreover, one can show that $X_{n,j}^P, j = 1, 2..., J$ is an approximation of the prediction distribution $f_{X_n|Y_{1:n-1}}(x_n|y_{1:n-1})$. After reweighting, $X_{n,j}^F, j = 1, 2..., J$ is an approximation of the filtering distribution $f_{X_n|Y_{1:n}}(x_n|y_{1:n})$, see Gordon et al. (1993); Smith and Gelfand (1992).

Another way of computing the likelihood is by noticing:

Prediction Formula: $\quad f_{X_n|Y_{1:n-1}}(x_n|y_{1:n-1}) = \int f_{X_n|X_{n-1}}(x_n|x_{n-1}) f_{X_{n-1}|Y_{1:n-1}}(x_{n-1}|y_{1:n-1}) dx_{n-1}$ (4)

Filtering Formula: $\quad f_{X_n|Y_{1:n}}(x_n|y_{1:n}) = \dfrac{f_{Y_n|X_n}(y_n|x_n) f_{X_n|Y_{1:n-1}}(x_n|y_{1:n-1})}{f_{Y_n|Y_{1:n-1}}(y_n|y_{1:n-1})}$ (5)

Conditional Likelihood: $\quad f_{Y_n|Y_{1:n-1}}(y_n|y_{1:n-1}) = \int f_{Y_n|X_n}(y_n|x_n) f_{X_n|Y_{1:n-1}}(x_n|y_{1:n-1}) dx_n$ (6)

When $f_{X_n|Y_{1:n}}(x_n \mid y_{1:n})$ and $f_{X_n|Y_{1:n-1}}(x_n \mid y_{1:n-1})$ are known functions of n, the likelihood can also be computed for all n, as these probabilities are inherently related to each other.

Since

$$f_{X_{n-1}|Y_{1:n-1}}(x_{n-1}|y_{1:n-1}) \xrightarrow{\text{prediction}} f_{X_n|Y_{1:n-1}}(x_n|y_{1:n-1}) \xrightarrow{\text{update}} f_{X_n|Y_{1:n}}(x_n|y_{1:n})$$

However, the evaluation of those integrals is almost impossible due to their complex and high-dimensional nature. Thus obtaining approximations for this distribution is more welcome (Doucet et al., 2001).

2.3 Poisson Approximate Likelihood (PAL)

Denote $p_{X_n|Y_{1:n-1}}(x_n|y_{1:n-1})$ as the prediction density; $p_{X_n|Y_{1:n}}(x_n|y_{1:n})$ as the filtering density. Suppose $\{X_n : n \in \mathbb{N}\}$ and $\{Y_n : n \in \mathbb{N}\}$ are discrete random variables, with $\mathcal{X} = \mathbb{N}^m, \mathcal{Y} = \mathbb{N}$, and $m \in \mathbb{N}$ is the number of compartments, by very minimum algebra, we have:

Prediction Formula: $\quad p_{X_n|Y_{1:n-1}}(x_n|y_{1:n-1}) = \displaystyle\sum_{x_{n-1} \in \mathbb{N}_0^m} p_{X_n|X_{n-1}}(x_n|x_{n-1}) p_{X_{n-1}|Y_{1:n-1}}(x_{n-1}|y_{1:n-1})$

Filtering Formula: $\quad p_{X_n|Y_{1:n}}(x_n|y_{1:n}) = \dfrac{p_{Y_n|X_n}(y_n|x_n) p_{X_n|Y_{1:n-1}}(x_n|y_{1:n-1})}{p_{Y_n|Y_{1:n-1}}(y_n|y_{1:n-1})}$

Conditional Likelihood: $\quad p_{Y_n|Y_{1:n-1}}(y_n|y_{1:n-1}) = \displaystyle\sum_{x_n \in \mathbb{N}_0^m} f_{Y_n|X_n}(y_n|x_n) p_{X_n|Y_{1:n-1}}(x_n|y_{1:n-1})$

which is a discrete representation of Eq. (4)-(6).

Given $p_{X_0}(x_0)$, $p_{X_n|X_{n-1}}(x_n|x_{n-1})$, and $p_{Y_n|X_n}(y_n|x_n)$ for $n \geq 1$, we can approximate $p_{X_n|Y_{1:n-1}}(x_n|y_{1:n-1})$, $p_{X_n|Y_{1:n}}(x_n|y_{1:n})$, and $p_{Y_n|Y_{1:n-1}}(y_n|y_{1:n-1})$ at any time $n \geq 1$ and then use the above quantities to update $p_{X_n|X_{n-1}}(x_n|x_{n-1}), p_{X_n|Y_{1:n}}(x_n|y_{1:n})$, and $p_{X_n|Y_{1:n}}(x_n|y_{1:n})$ at the next given time $n + 1$. The Poisson Approximate Likelihood (PAL) method assumes and approximates them by a vector Poisson approximation.

1. In the latent process model:

    (a) $\boldsymbol{\delta_t}$ be the death rate vector;

    (b) $\boldsymbol{\alpha_t}$ be the birth rate vector of m compartments;

    (c) $\boldsymbol{K_{t,\eta}}$ be the row stochastic matrices;

    Then the one-step transition of the latent process can be described as:

    $\boldsymbol{x_t} = \boldsymbol{\tilde{x}_t} + \boldsymbol{\hat{x}_t}$, where $\tilde{\boldsymbol{x}}$ is obtained by $\bar{\boldsymbol{x}}_{t-1} \sim Binom(\boldsymbol{x}_{t-1}, \boldsymbol{1} - \boldsymbol{\delta_t})$ and $\tilde{\boldsymbol{x}}_t$ is equal to the one-step transition of $\bar{\boldsymbol{x}}_{t-1}$ according to the transition kernel $\boldsymbol{K_{t-1,\eta}}$. And $\hat{\boldsymbol{x}}_t \sim Pois(\boldsymbol{\alpha_t})$ is the importing birth, which is considered as stochastic noise.

2. In the observation model:

    (a) $\boldsymbol{q_t} \in [0,1]^m$ be the observing probability, where m is the number of compartments;

    (b) $\boldsymbol{G_t}$ be the misreporting kernel, where non-zero off-diagonal entries indicate misreporting.

    (c) $\boldsymbol{\kappa_t}$ be the observation error rate.

    Then $\boldsymbol{y_t} = \tilde{\boldsymbol{y}}_t + \hat{\boldsymbol{y}}_t$, where $\hat{\boldsymbol{y}}_t \sim Binom(\boldsymbol{x_t}, \boldsymbol{q_t})$. And $\hat{\boldsymbol{y}}_t \sim Pois(\boldsymbol{\kappa_t})$ is the observation error.

A novel method suggested by Whitehouse et al. (2023) is that the conditional likelihood of the above compartmental model can be approximated by Poisson densities by the following:

If we assume $p_{X_0}(\boldsymbol{x}_0) \approx \text{Pois}(\boldsymbol{\lambda}_0)$, where $\boldsymbol{\lambda}_0$ is a parameter that characterizes the initial distribution of the numbers of individuals in each compartment.

Then for $t \geq 1$:

1. $p(\boldsymbol{x}_t|\boldsymbol{y}_{1:t-1}) \approx \text{Pois}(\boldsymbol{\lambda}_t)$, where $\boldsymbol{\lambda}_t := (\bar{\boldsymbol{\lambda}}_{t-1} \odot \boldsymbol{\delta}_t)^{\top} \boldsymbol{K}_{t,\eta(\bar{\boldsymbol{\lambda}}_{t-1} \odot \boldsymbol{\delta}_t)} + \boldsymbol{\alpha}_t$

2. $p(\mathbf{x}_t \mid \mathbf{y}_{1:t}) \approx \text{Pois}(\bar{\boldsymbol{\lambda}}_t), \quad \bar{\boldsymbol{\lambda}}_t := \left[\mathbf{1}_m - \mathbf{q}_t + \left(\left\{ \mathbf{y}_t^{\top} \oslash \left[(\mathbf{q}_t \odot \boldsymbol{\lambda}_t)^{\top} \mathbf{G}_t + \boldsymbol{\kappa}_t^{\top} \right] \right\} \left[(\mathbf{1}_m \otimes \mathbf{q}_t) \odot \mathbf{G}_t^{\top} \right] \right)^{\top} \right] \odot \boldsymbol{\lambda}_t$

3. $p(\mathbf{y}_t|\mathbf{y}_{1:t-1}) \approx \text{Pois}(\boldsymbol{\mu}_t)$, where $\boldsymbol{\mu}_t = \left[(\boldsymbol{\lambda}_t \odot \mathbf{q}_t)^{\top} \mathbf{G}_t \right]^{\top} + \boldsymbol{\kappa}_t$.

Therefore, the algorithm of the approximated likelihood is:

Algorithm 2 Poisson Approximate Likelihood (PAL)

   Initialize: $\bar{\lambda}_0 \leftarrow \lambda_0$.

1: **for** t in $1:N$ **do**

2:    $\lambda_t \leftarrow \left[(\bar{\lambda}_{t-1} \odot \delta_t)^{\top} \mathbf{K}_{t,\eta(\bar{\lambda}_{t-1} \odot \delta_t)} \right]^{\top} + \boldsymbol{\alpha}_t$

3:    $\bar{\lambda}_t \leftarrow \left[\mathbf{1}_m - \mathbf{q}_t + \left(\left\{ \mathbf{y}_t^{\top} \oslash \left[(\mathbf{q}_t \odot \lambda_t)^{\top} \mathbf{G}_t + \boldsymbol{\kappa}_t^{\top} \right] \right\} \left[(\mathbf{1}_m \otimes \mathbf{q}_t) \odot \mathbf{G}_t^{\top} \right] \right)^{\top} \right] \odot \lambda_t$

4:    $\boldsymbol{\mu}_t \leftarrow \left[(\lambda_t \odot \mathrm{q}_t)^{\top} \mathbf{G}_t \right]^{\top} + \boldsymbol{\kappa}_t$

5:    $\ell(\mathbf{y}_t \mid \mathbf{y}_{1:t-1}) \leftarrow -\boldsymbol{\mu}_t^{\top} \mathbf{1}_m + \mathbf{y}_t^{\top} \log(\boldsymbol{\mu}_t) - \mathbf{1}_m^{\top} \log(\mathbf{y}_t!)$

6: **end for**

To account for overdispersion within the model, Whitehouse et al. (2023)'s approach entails the introduction of a prior distribution over the parameters. If we separate parameters into two parts i.e. $\theta = [\vartheta, \bar{\theta}_{1:t}]$ where ϑ are either fixed or to be estimated and $\bar{\theta}_{1:t}$ are to be integrated out and $\bar{\theta}_{1:t} \sim f(\cdot|\phi)$, the prior distribution for $\bar{\theta}_{1:t}$, where ϕ are hyperparameters. These parameters $\bar{\theta}_{1:t}$ are then integrated out by importance sampling techniques with proposal distribution $\pi(\cdot|\cdot)$, yielding a marginal distribution of the data that reflects the overdispersion. The procedure is delineated in the following algorithm:

Algorithm 3 PALSMC

Input: Proposal distribution $\pi(\cdot|\cdot)$, number of particles J, parameter $[\vartheta, \phi]$
Initialize: $\bar{\lambda}_0^{(i)} \leftarrow \lambda_0$, for $i = 1, 2, ..., J$.

1: **for** t in $1 : N$ **do**
2:     **for** $i = 1, 2, ..., J$ **do**
3:         $\bar{\theta}_{1:t}^{(i)} \sim \pi(\cdot|\bar{\theta}_{1:t-1}^{(i)}, \lambda_{t-1}^{(\bar{i})}, y_{1:t})$.
4:         $\boldsymbol{\alpha}_t \leftarrow \boldsymbol{\alpha}_t(\vartheta, \bar{\theta}_t^{(i)}), \mathbf{K}_{t,\boldsymbol{\eta}} \leftarrow \mathbf{K}_{t,\boldsymbol{\eta}}(\vartheta, \bar{\theta}_t^{(i)}), \boldsymbol{\delta}_t \leftarrow \boldsymbol{\delta}_t(\vartheta, \bar{\theta}_t^{(i)}), \boldsymbol{q}_t \leftarrow \boldsymbol{q}_t(\vartheta, \bar{\theta}_t^{(i)}),$
        $\boldsymbol{\kappa}_t \leftarrow \boldsymbol{\kappa}_t(\vartheta, \bar{\theta}_t^{(i)}), \boldsymbol{G}_t \leftarrow \boldsymbol{G}_t(\vartheta, \bar{\theta}_t^{(i)})$
5:         $\lambda_t \leftarrow \left[\left(\bar{\lambda}_{t-1} \odot \delta_t\right)^{\top} \mathbf{K}_{t,\boldsymbol{\eta}(\bar{\lambda}_{t-1} \odot \delta_t)} \right]^{\top} + \boldsymbol{\alpha}_t$
6:         $\bar{\lambda}_t \leftarrow \left[\mathbf{1}_m - \mathbf{q}_t + \left(\left\{ \mathbf{y}_t^{\top} \oslash \left[\left(\mathbf{q}_t \odot \lambda_t\right)^{\top} \mathbf{G}_t + \boldsymbol{\kappa}_t^{\top} \right] \right\} \left[\left(\mathbf{1}_m \otimes \mathbf{q}_t\right) \odot \mathbf{G}_t^{\top} \right] \right)^{\top} \right] \odot \lambda_t$
7:         $\boldsymbol{\mu}_t \leftarrow \left[\left(\lambda_t \odot \mathrm{q}_t\right)^{\top} \mathbf{G}_t \right]^{\top} + \boldsymbol{\kappa}_t$
8:         $\ell\left(\mathbf{y}_t \mid \mathbf{y}_{1:t-1}, \vartheta, \bar{\theta}_{1:t}^{(i)}\right) \leftarrow -\boldsymbol{\mu}_t^{\top} \mathbf{1}_m + \mathbf{y}_t^{\top} \log\left(\boldsymbol{\mu}_t\right) - \mathbf{1}_m^{\top} \log\left(\mathbf{y}_t!\right)$
9:         $\log w_t^{(i)} \leftarrow \ell\left(\mathbf{y}_t \mid \mathbf{y}_{1:t-1}, \vartheta, \bar{\theta}_{1:t}^{(i)}\right) + \log f(\bar{\theta}_t^{(i)}|\bar{\theta}_{1:t-1}^{(i)}, \phi) - \log \pi(\bar{\theta}_t^{(i)}|\bar{\theta}_{1:t-1}^{(i)}, \bar{\lambda}_t^{(i)})$
10:     **end for**
11:     $\hat{\ell}\left(\mathbf{y}_t \mid \mathbf{y}_{1:t-1}, \vartheta, \phi\right) \leftarrow \log\left(\frac{1}{J} \sum_{i=1}^{J} w_t^{(i)}\right)$
12:     $\bar{w}_t^{(i)} \leftarrow \frac{w_t^{(i)}}{\sum_{j=1}^{J} w_t^{(j)}}$
13:     Resample $\{\bar{\theta}_{1:t}^{(i)}, \bar{\lambda}_t^{(i)}\}$ according to weights $\{\bar{w}_t^{(i)}\}_{i=1}^{J}$
14: **end for**

---

3   Model: A Case Study with Rotavirus in Germany: 2001-2008

Table 1: Comparison of Model & Method Performance (Whitehouse et al., 2023)

Model	AIC	Ave. comp. time
EqEq	98866.65	30 s
EqOv	15154.75	2 hr
OvOv	13778.08	3 hr
Stocks et al. (2018)	20134.38	11 hr

Model selection is crucial in epidemiology modeling, as it determines the appropriateness of a model in capturing the complexities of disease transmission dynamics. The Akaike Information Criterion (AIC) represents a mathematical approach used in the selection of models and the evaluation of their structural parsimony. The formula for this criterion is given by.

$$AIC = 2p - \ell(\hat{\theta})$$

where p is the number of parameters in the model and $\ell(\hat{\theta})$ is the maximized log-likelihood at the estimates.

When faced with multiple potential models for a dataset, the optimal choice is the model that yields the lowest AIC value. The AIC not only acknowledges the model's fit, evaluated by the likelihood function but also imposes a penalty that escalates with the number of parameters estimated within the model. This penalization is intended to deter overfitting, which is advantageous because augmenting the parameter count typically enhances the model's fit regardless of its true explanatory power.

Next, we introduce the POMP models used in the analysis, beginning with the latent process model, which follows a Markovian structure and includes nine compartments within $\mathcal{X}$.

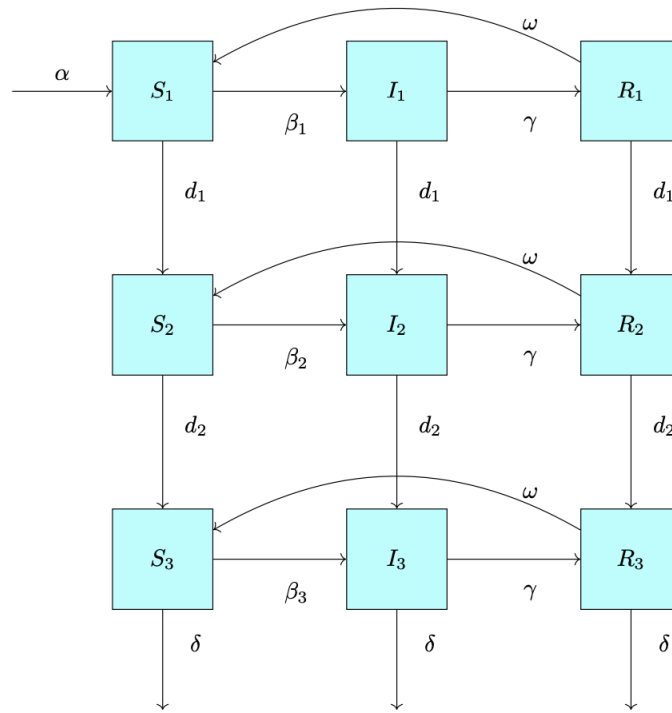## 3.1 Latent Process Compartmental Model: The SIRSIRSIR model

Figure 2: SIRSIRSIR model (Whitehouse et al., 2023)

Rotavirus disease primarily affects infants and young children, causing gastroenteritis in this age group, and is relatively rare in adults (Lambert et al., 2009). Nearly every child contracts rotavirus at least once by the age of five (CDC, 2020). Reinfection can occur since neither natural infection nor vaccination against rotavirus provides complete immunity against future infections (Stocks et al., 2018). Consequently, it is practical to categorize the population into three age groups, leading to a model with nine compartments in total: S_k, susceptible; I_k, infected; R_k, recovered, where $k = 1, 2, 3$ corresponds to the age classes 0-4, 5-59, and 60-99, respectively. The diagram of the latent model is shown in Figure 2, which is employed by both Whitehouse et al. (2023) and Stocks et al. (2018). The model is explained in detail subsequently.

The evolution of the full age-stratified rotavirus latent process model at time t is given by:

$$S_{1,t+1} = S_{1,t-1} + A_{1,t} + E_{1,t} - B_{1,t} - F_{1,t}^{(S)},$$

$$I_{1,t+1} = I_{1,t} + B_{1,t} - C_{1,t} - F_{1,t}^{(I)},$$

$$R_{1,t+1} = R_{1,t} + C_{1,t} - E_{1,t} - F_{1,t}^{(R)},$$

$$S_{2,t+1} = S_{2,t} + F_{1,t}^{(S)} + E_{2,t} - B_{2,t} - F_{2,t}^{(S)},$$

$$I_{2,t+1} = I_{2,t} + F_{1,t}^{(I)} + B_{2,t} - C_{2,t} - F_{2,t}^{(I)},$$

$$R_{2,t+1} = R_{2,t} + F_{1,t}^{(R)} + C_{2,t} - E_{2,t} - F_{2,t}^{(R)},$$

$$S_{3,t+1} = S_{3,t} + F_{2,t}^{(S)} + E_{3,t} - B_{3,t} - D_t^{(S)},$$

$$I_{3,t+1} = I_{3,t} + F_{2,t}^{(I)} + B_{3,t} - C_{3,t} - D_t^{(I)},$$

$$R_{3,t+1} = R_{3,t} + F_{2,t}^{(R)} + C_{3,t} - E_{3,t} - D_t^{(R)},$$

where at time t: $A_{1,t} \sim \text{Pois}(\alpha_t)$, for some $\alpha_t \in \mathbb{R}$ represents new births, which is chosen according to historical birth record data; $B_{\cdot,t}$ represents new infectives; $C_{\cdot,t}$ represents recovering individuals; $D_{\cdot,t} \sim \text{Binom}(t-1, 1-\delta)$ represents emigrating (dying) individuals; $E_{\cdot,t}$ represents individuals experiencing waning immunity; and $F_{\cdot,t}$ represents aging individuals.

$$\begin{bmatrix} B_{1,t} \\ F_{1,t}^{(S)} \\ S_{1,t} - B_{1,t} - F_{1,t}^{(S)} \end{bmatrix} \sim \text{Mult}\left(S_{1,t} , \begin{bmatrix} p_{1,t} \\ 1 - e^{-hd1} \\ e^{-hd1} - p_{1,t} \end{bmatrix} \right)$$

$$\begin{bmatrix} C_{1,t} \\ F_{1,t}^{(I)} \\ I_{1,t} - C_{1,t} - F_{1,t}^{(I)} \end{bmatrix} \sim \text{Mult}\left(I_{1,t} , \begin{bmatrix} 1 - e^{-h\gamma} \\ 1 - e^{-hd1} \\ e^{-h\gamma} + e^{-hd1} - 1 \end{bmatrix} \right)$$

$$\begin{bmatrix} E_{1,t} \\ F_{1,t}^{(R)} \\ R_{1,t} - E_{1,t} - F_{1,t}^{(R)} \end{bmatrix} \sim \text{Mult}\left(R_{1,t} , \begin{bmatrix} 1 - e^{-h\omega} \\ 1 - e^{-hd1} \\ e^{-h\omega} + e^{-hd1} - 1 \end{bmatrix} \right)$$

$$\begin{bmatrix} B_{2,t} \\ F_{2,t}^{(S)} \\ S_{2,t} - B_{2,t} - F_{2,t}^{(S)} \end{bmatrix} \sim \text{Mult}\left(S_{2,t} , \begin{bmatrix} p_{2,t} \\ 1 - e^{-hd2} \\ e^{-hd2} - p_{2,t} \end{bmatrix} \right)$$

$$\begin{bmatrix} C_{2,t} \\ F_{2,t}^{(I)} \\ I_{2,t} - C_{2,t} - F_{2,t}^{(I)} \end{bmatrix} \sim \text{Mult}\left(I_{2,t} , \begin{bmatrix} 1 - e^{-h\gamma} \\ 1 - e^{-hd2} \\ e^{-h\gamma} + e^{-hd2} - 1 \end{bmatrix} \right)$$

$$\begin{bmatrix} E_{2,t} \\ F_{2,t}^{(R)} \\ R_{2,t} - E_{2,t} - F_{2,t}^{(R)} \end{bmatrix} \sim \text{Mult}\left(R_{2,t} \;,\; \begin{bmatrix} 1 - e^{-h\omega} \\ 1 - e^{-hd2} \\ e^{-h\omega} + e^{-hd2} - 1 \end{bmatrix} \right)$$

$$\begin{bmatrix} B_{3,t} \\ S_{3,t} - D_t^{(S)} - B_{3,t} \end{bmatrix} \sim \text{Mult}\left(S_{3,t} - D_t^{(S)} \;,\; \begin{bmatrix} p_{3,t} \\ 1 - p_{3,t} \end{bmatrix} \right)$$

$$\begin{bmatrix} C_{3,t} \\ I_{2,t} - D_t^{(I)} - C_{2,t} \end{bmatrix} \sim \text{Mult}\left(I_{3,t} - D_t^{(I)} \;,\; \begin{bmatrix} 1 - e^{-h\gamma} \\ e^{-h\gamma} \end{bmatrix} \right)$$

$$\begin{bmatrix} E_{3,t} \\ R_{2,t} - D_t^{(R)} - E_{2,t} \end{bmatrix} \sim \text{Mult}\left(R_{3,t} - D_t^{(R)} \;,\; \begin{bmatrix} 1 - e^{-h\omega} \\ e^{-h\omega} \end{bmatrix} \right)$$

$$\mathbf{K}_{t,\boldsymbol{\eta}} = \begin{bmatrix}
-e^{-hd1} & p_{1,t} & 0 & 1-e^{-hd1} & 0 & 0 & 0 & 0 & 0 \\
0 & -e^{-h\gamma}+e^{-hd1}-1 & 1-e^{-h\gamma} & 1-e^{-hd1} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1-e^{-h\omega} & 0 & e^{-h\omega}+e^{-hd1}-1 & 1-e^{-hd1} & 0 & 0 & 0 \\
0 & 0 & 0 & -e^{-hd2} & p_{2,t} & 0 & 1-e^{-hd2} & 0 & 0 \\
0 & 0 & 0 & 0 & -e^{-h\gamma}+e^{-hd2}-1 & 1-e^{-h\gamma} & 1-e^{-hd2} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1-e^{-h\omega} & 0 & e^{-h\omega}+e^{-hd2}-1 & 1-e^{-hd2} \\
0 & 0 & 0 & 0 & 0 & 0 & -e^{-hd3} & p_{3,t} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -e^{-h\gamma}+e^{-hd3}-1 & 1-e^{-h\gamma} \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1-e^{-h\omega}
\end{bmatrix}$$

where Mult denotes multinomial distribution. For models EqEq and EqOv we have $p_{k,t} = 1 - \exp\left\{ -\beta_k \frac{I_t}{n} \chi_t \right\}$ for $k = 1,2,3$, and for model OvOv we have $p_{k,t} = 1 - \exp\left\{ -\beta_k \frac{\tilde{I}_t}{n} \chi_t \xi_t \right\}$ for $k = 1,2,3$, $\chi_t = 1 + \rho \cos(2\pi t / p + \phi)$, where $p = 52$ is the fixed period corresponding to the annual outbreak of rotavirus.

3.2 Observation Model

The observation model is defined as follows:

$$[Y_{kt} \mid X_{kt} = (S_{kt}, I_{kt}, R_{kt})] = [Y_{kt} \mid I_{kt}], \quad k = 1,2,3 \quad t = 1,2,...,416$$

$$\sim p(y_{kt} \mid x_{kt})$$

This model covers three age groups, $k = 1,2,3$, and specifies that the observable variable Y_{kt} is determined solely by I_{kt}, which represents the number of infected individuals at time t. The time variable t spans from week 0 to week 416, corresponding to an eight-year period from 2001 to 2008. This assumption is practical because most reported cases are confirmed either by laboratory tests or by clinical diagnosis in hospitals of infected individuals. The observed counts are modeled using either a binomial distribution, as assumed by Whitehouse et al. (2023), or a Poisson or Negative Binomial distribution, as assumed by Stocks et al. (2018). Both approaches use the conditional mean equal to the product of

the number of infectious individuals and the reporting rate q_t. This model reflects the reality that researchers can only collect data on reported cases, which represent only a subset of actual infections due to underreporting.

For the observation models of Stocks et al. (2018) and Whitehouse et al. (2023) are defined as follows:

1. (Stocks et al., 2018):

$$Y_{kt} \sim \text{Pois}(q_t \cdot H_{kt}) \quad \text{or} \quad Y_{kt} \sim \text{NegBinom}(q_t \cdot H_{kt}, \frac{1}{\theta})$$

In the model proposed by Stocks et al. (2018), the number of observed cases Y_{kt} is modeled to follow either a Poisson distribution or a Negative Binomial distribution, to account for potential overdispersion in the observation model. The expected value of these distributions is the product of the reporting rate q_t and the number of infected individuals H_{kt}. H_{kt} is modeled as an accumulated variable, in the sense that it's the total number of transition from I_k to R_k within a week, corresponding to the weekly data we use. The Negative Binomial distribution includes a dispersion parameter θ to manage overdispersion relative to the Poisson model.

However, Stocks et al. (2018) assumed a non-underreporting observation model, implying $q_t = 1$, which means the data they analyzed were effectively scaled up. This assumption led to an increase of 3200 log-units in the log-likelihood, as reported by Whitehouse et al. (2023) and illustrated in Table 1.

2. (Whitehouse et al., 2023):

$$Y_{kt} \sim \text{Binom}(H_{kt}, q_t)$$

In Whitehouse et al. (2023)'s model, Y_{kt} follows a Binomial distribution, where H_{kt} represents the trials and q_t the success probability in each trial, reflecting the proportion of infections that are actually detected and reported.

Focusing on Whitehouse et al. (2023)'s modeling, we consider three models:

1. EqEq: a fully equi-dispersed model, in which $q_{k,r} = 0.07 = \mu_q \in [0,1]$, and μ_q is assumed to be known;

2. EqOv: an equi-dispersed latent compartmental model and an over-dispersed observation model, the same as EqEq except that $q_{k,r} \overset{\text{iid}}{\sim} \mathcal{N}(\mu_q, \sigma_q^2)_{\geq 0, \leq 1}$, where $\sigma_q^2 > 0$ is to be estimated;

3. OvOv: over-dispersion in both the latent and observation models, the same as EqOv except that we augment χ_t to $\chi_t \xi_{t,r}$, where for $r \geq 1$, $\xi_{t,r} \overset{\text{iid}}{\sim} \text{Gamma}(\sigma_\xi^2, \sigma_\xi^2)$ are multiplicative disturbances with mean 1 and $\sigma_\xi^2 > 0$ is to be estimated.

The table of parameters can be found below. Note that some of them are assumed fixed while the rest are to be estimated.

Table 2: Full-list of parameters

Parameter	Explanation	Value	Unit
N	Population size	82372825	individuals
α	Individual birth rate	$1/(52 \cdot 78.9)$	week^{-1}
d_1	Aging from age group 1 (age 0-4) to 2 (age 5-59)	$1/(52 \cdot 5)$	week^{-1}
d_2	Aging from age group 2 (age 5-59) to 3 (age 60+)	$1/(52 \cdot 55)$	week^{-1}
δ	Death rate (from age group 3 (age 60+))	$1/(52 \cdot 18.9)$	week^{-1}
β_k	Susceptibility of age group $k \in \{1, 2, 3\}$	To be estimated	individuals/week
ω	Immunity waning rate	$1/(1 \cdot 52)$	week^{-1}
γ	Recovery rate	1	week^{-1}
ϕ	Phase shift of the seasonal forcing	To be estimated	1
ρ	Amplitude of the seasonal forcing	To be estimated	1
$p_{k,t}$	Force of infection of age group $k \in \{1, 2, 3\}$	Defined above	week^{-1}
$\chi(t)$	Seasonal forcing function	Defined above	1
σ_q^2	Overdispersion parameter of reporting rates	To be estimated	1
σ_ξ	Scale parameter of the Γ-noise	To be estimated	1

4   Results

Whitehouse et al. (2023) claimed that PAL is the first likelihood-based filtering method that can be applied to a broad range of POMP models with the finite-population assumption and consistent results, and requires much less computational time than SMC. 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 model 1000 times at the MLE shown in Table 3 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. These results are presented in Appendix A.

4.1   Key Factors Contributing to the Change of AIC in Table 1

From Table 1, it is evident that the OvOv model yields the lowest AIC, making it particularly noteworthy. Since one of Stocks et al. (2018)'s main purposes was to do model selection, therefore it's exciting to see Whitehouse et al. (2023) obtaining a much better model. The overdispersion model (OvOv) is often the best model in terms of likelihood, primarily because it allows for greater variation in the model. By incorporating more parameters, the OvOv model can capture subtleties and complexities in the dataset that simpler models may overlook, leading to a more accurate fit and higher likelihood values. Whitehouse et al. (2023) claimed that it is the innovative observation model that contributed to an increase

Table 3: Maximum likelihood estimation of parameters by PAL for three models (White-house et al., 2023), and by mif2 algorithm (King et al., 2016) for the OvOv model.

Parameter	EqEq	EqOv	OvOv	OvOv (**pomp**)[*]
β_1	12.15	12.74	11.48	10.30
β_2	0.22	0.21	0.25	0.21
β_3	0.34	0.31	0.35	0.86
ϕ	0.017	0.14	0.14	0.03
ρ	0.022	0.19	0.16	0.14
σ_q^2	n/a	0.042	0.021	0.018
σ_ξ	n/a	n/a	66.89	14.85
S_{10}		3876549[*]		5885201
I_{10}		30351		5041
R_{10}		1315221		721006
S_{20}		57139612		44546887
I_{20}		871		22
R_{20}		302852		26360811
S_{30}		19573727		4788420
I_{30}		2550		131
R_{30}		131092		65306
AIC	98866.65	15154.75	13778.08	**13674.55**

[*] 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.

in log-likelihood by 3200 log-likelihood units. However, after comparing the results, we found that the increase in likelihood when implementing the OvOv model is attributable to the datasets used by White-house et al. (2023) and Stocks et al. (2018), where Stocks et al. (2018)'s dataset has no underreporting. Therefore Whitehouse et al. (2023)'s dataset can be viewed as a scaled-down version of Stocks et al. (2018)'s dataset by the reporting rate, which is roughly 0.07. Within the framework of the same model, scaling the data downwards by a factor results in a corresponding rise in likelihood. This scaling effect accounts for the discrepancy observed in the increase of 3200 log-likelihood units.

4.2 PAL versus SMC: When the Model is Misspecified

Our results demonstrate how SMC is sensitive to model misspecification, which consequently leads to a low likelihood estimate at the MLE, as shown in Table 4 by fitting the real German rotavirus dataset into the model.

Table 4: 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	-6892.168	-7200.866

Table 4 indicates that the results from the two methods are quite similar but SMC obtained a bit lower log-likelihood estimate than PAL. One potential explanation is the model misspecification, leading to a sensitive detection of likelihood shortfalls for SMC. Since consistent datasets, when faced with

model misspecification due to outliers in the data, challenge the SMC approach, leading to low weights being assigned to almost all particles (recall Algorithm 1), as the latent process model struggles to accommodate outliers and is misspecified for such data points, which results in low conditional likelihood at that specific data point. Moreover, for SMC, a situation where a few particles acquire a high weight while the majority are assigned negligible weights can lead to an imbalance in the resampling process. This scenario might result in these few heavily weighted particles being resampled repeatedly, thereby dominating the sample set. If, following this, the subsequent simulations are unable to track the model accurately, it is regarded as an SMC failure. However, such an occurrence was not observed in our analysis, indicating that the SMC method functioned without encountering this specific type of failure.

To examine the model misspecification at potential time points, as depicted in Figure 3, the majority of significant drops of the SMC occurred at time ($t = 1, 2, 3, 11, 81, 194, 325 - 333$). The significant drops observed initially, specifically at times $t = 1, 2, 3$, may suggest an improper initial distribution from which the first set of particles was simulated. At times $t = 11, 81, 194$ and $t = 325 - 333$, the shortfall approximated $50 - 100$ log-units. These failures can be attributed to outliers that the model cannot explain, as shown in Appendix B. This suggests model misspecification, particularly in the latent process, which fails to generate suitable particles capable of producing a high likelihood of observing the true reported cases from the data. Subsequent issues observed at later times could indicate problems caused by fixed seasonality, which cannot be compensated for by adding more stochastic elements.
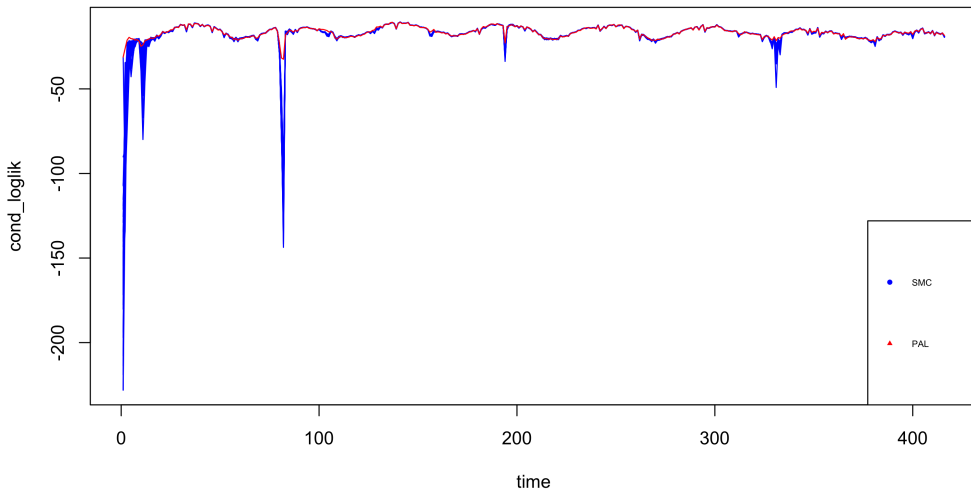

Figure 3: Conditional log-likelihoods computed using two methods for the real rotavirus dataset, with the SMC method replicated 36 times due to its high variances. The blue line represents overlapping results from 36 SMC computations, while the red line is derived from PAL. The main sources of likelihood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

4.3 PAL versus SMC: When the Model is Correctly Specified

Here, we utilize 100 simulated datasets from the OvOv model with parameters set at the MLE, thereby eliminating model misspecification. We then examine the likelihood values computed by PAL and SMC.

Table 5: Average log-likelihood at the MLE at the 100 simulated datasets from OvOv computed by two filtering methods, employed 36 replicates and 50,000 particles

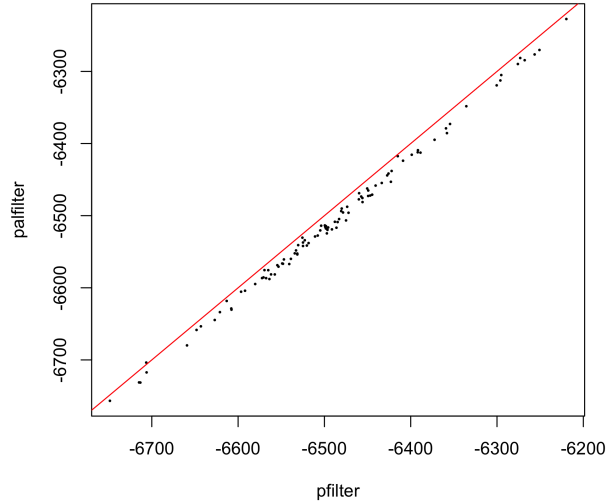|	PAL	SMC
OvOv	-6231.97	-6224.245

Figure 4: Log-likelihoods computed using two filtering methods for 100 randomly simulated datasets at the MLE of the OvOv model. The red line is the line $Lik(pfilter) = Lik(palfilter)$. We can see on the simulated data, SMC gives consistently higher likelihood estimates than PAL.

When the model is correctly specified, however, the log-likelihood values computed by PAL, as shown in Table 5 and Figure 4, consistently exhibit a shortfall. This shortfall highlights the approximative nature of the PAL method compared to the more accurate SMC filter. It is posited that the model yielding the highest likelihood is, in theory, the true model, a concept we will explore in further detail in the discussion section 5.

4.4 Iterated Filtering Maximization of the Likelihood

Besides proposing a novel filtering algorithm, by applying PAL to the model for rotavirus, Whitehouse et al. (2023) claimed they have also identified a superior model, as a much lower AIC represented in Table 1 indicates and the failure from getting a lower likelihood in Section 4.2. However, as demonstrated in 4.3, we recognize that SMC, functioning as an ideal filter, ought to yield a higher likelihood estimation when the model is correct. Notably, Section 4.2 reveals that when the model confronts actual rotavirus data, it encounters misspecification issues, particularly at times $t = 1, 2, 3, 11, 81, 194, 325 - 333$, leading to shortfalls in likelihood. Appendix B discusses this phenomenon, attributing it to certain data points that the model fails to adequately represent. This issue is especially pronounced at the initial time points. Particularly, in the initial time points $t = 1, 2, 3$, the likelihood deviates from PAL estimates for about more than 100 log-likelihood units, suggesting an improper choice of initial distribution that might favor the PAL estimate. Consequently, we have opted to estimate the initial distribution $\lambda_0 = (S_{10}, I_{10}, R_{10}, S_{20}, I_{20}, R_{20}, S_{30}, I_{30}, R_{30})$, where $S_{10}, I_{10}, R_{10}, S_{20}, I_{20}, R_{20}, S_{30}, I_{30}$, and R_{30} represent the

number of individuals in each compartment at time $t = 0$. The results of this new estimation can be found in Table 3. Importantly, the maximization of the likelihood conducted via the Iterated Filtering (mif2) algorithm (King et al., 2016) procures a higher likelihood at the estimated MLE than that estimated by PAL, and a different yet comparable set of estimates is obtained. In terms of parameters pertinent to epidemiological interests, i.e. the first 7 rows in Table 3, we observe a lower force of infection parameter for the first age group, β_1, alongside a higher force of infection for the third age group, β_3. Additionally, we detect variations in the phase parameter ϕ, which directly affects the seasonal time points, as well as in the stochastic parameter σ_ξ and the initial compartmental distributions. The modifications to the initial distributions aim to rectify inadequate initialization, while the other adjustments address model misspecifications identified in Section 4.2. Changes in the remaining parameter set are negligible. These findings indicate that when model misspecification is reduced to a minimal level, both PAL and SMC approaches tend to provide similar likelihood estimates, and the Iterated Filtering algorithm is capable of generating a superior maximum likelihood estimation than PAL equipped with the gradient ascent algorithm. Moreover, the results produced by SMC are more reliable due to the exact nature of the filter, despite PAL's computational efficiency.

5   Discussion and Conclusion

This paper examines two filtering approaches, with PAL specifically designed for certain types of compartmental POMP models. However, both methods have their limitations. Although this was not observed in our study, SMC is sensitive to model misspecification, which can lead to filtering failures. Conversely, PAL may experience a shortfall in likelihood estimation when the model is correctly specified.

PAL is an approximate filtering method used in scenarios where SMC is computationally prohibitive. It approximates the posterior distribution by a Poisson approximation. Due to its approximative nature, it does not perfectly capture the true posterior distribution. SMC, however, often referred to as the "perfect filter," uses a large set of particles and sophisticated resampling techniques to theoretically accurately approximate the posterior distribution. It is computationally more intensive than PAL but tends to provide a satisfying approximation, especially in complex models that are not compatible with PAL in general. When PAL and SMC are applied to a well-specified model meaning the model with the latent process model being correctly represented, SMC outperforms PAL because it better captures the nuances of the posterior distribution. This is why SMC might show a "likelihood shortfall" compared to PAL when the model is correctly specified. This is due to the approximation error of PAL can't be improved arbitrarily in general (Wills and Schön, 2023), and likelihood being a proper scoring rule.

The concept of likelihood as a proper scoring rule (Gneiting et al., 2007; Gneiting and Raftery, 2007) in statistical models implies that using likelihood-based methods will naturally align with the true model of the data, assuming the model structure correctly represents the underlying process. When the likelihood function is used as a scoring rule, it is considered proper because maximizing the likelihood leads

to parameter estimates that converge to their true values under large sample conditions, provided the model is specified correctly. A scoring rule is "strictly proper" if a forecaster maximizes the expected score by issuing a probabilistic forecast that exactly matches the true distribution of the outcomes. In other words, for a scoring rule to be strictly proper, the best strategy for the forecaster is to report the true probability distribution of the events they are forecasting. This promotes honest and accurate forecasting. This provides a theoretical explanation of why SMC outperforms PAL when the model is correct.

The "likelihood shortfall" in PAL when compared to SMC, particularly when the model is correct, highlights the trade-off between computational efficiency and accuracy in statistical approximations. While PAL may be faster and more computationally feasible in many scenarios, it sacrifices some accuracy, which becomes apparent when compared to the more exact SMC method, especially under the scrutiny of proper scoring rules as discussed in the context.

6   Acknowledgement

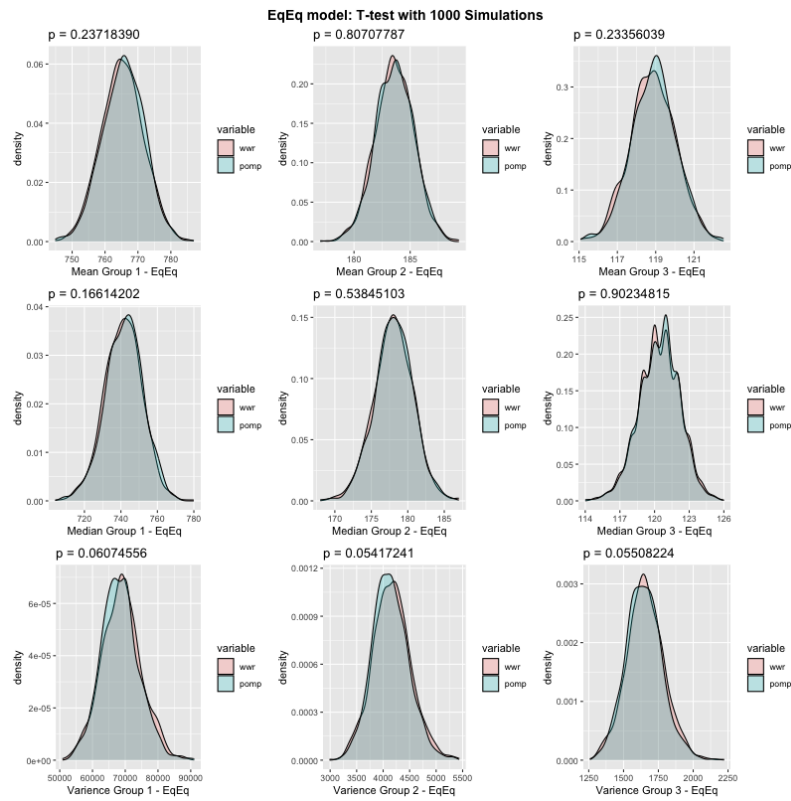# A   Consistency of Models Built in pomp and Whitehouse et al. (2023)

Figure 5: *Density plot and t-test of mean, median, and variance of all three age-specific time series of 1000 Simulations from EqEq model built in WWR code and **pomp***
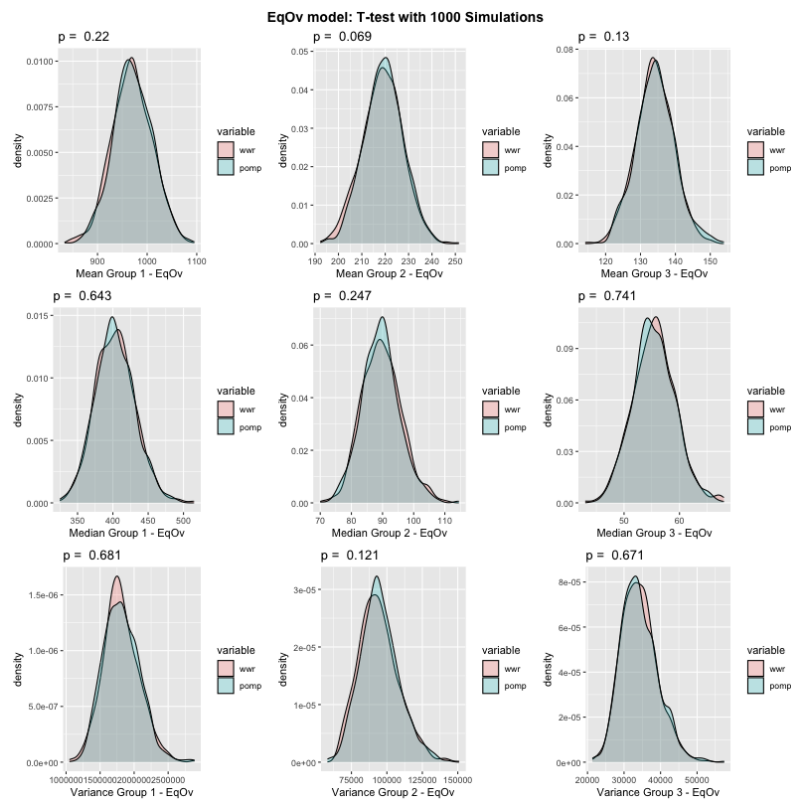

Figure 6: *Density plot and t-test of mean, median, and variance of all three age-specific time series of 1000 Simulations from EqOv model built in WWR code and **pomp***
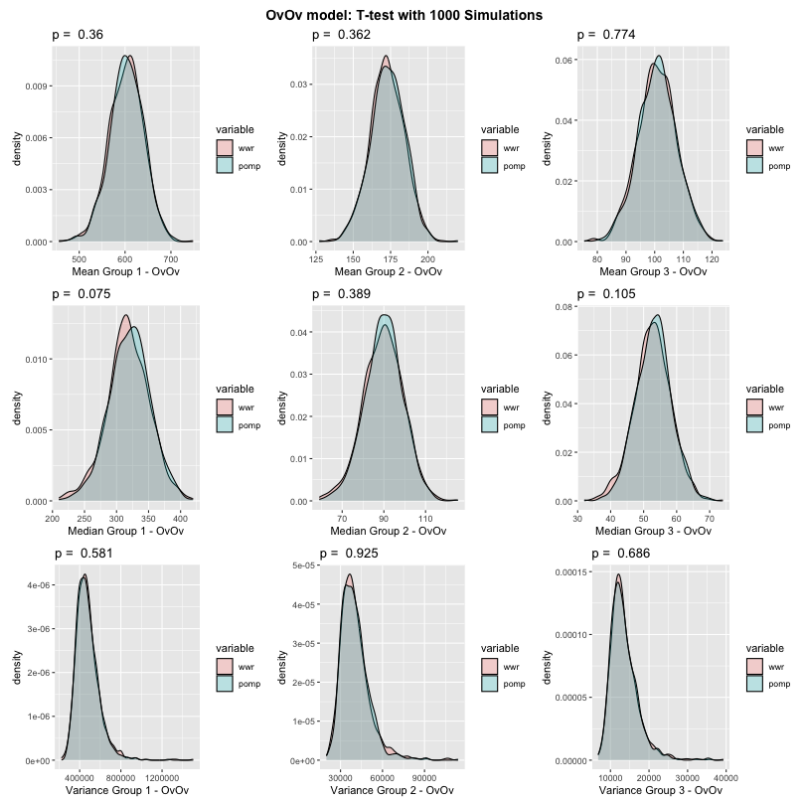
*Figure 7: Density plot and t-test of mean, median, and variance of all three age-specific time series of 1000 Simulations from OvOv model built in WWR code and **pomp***
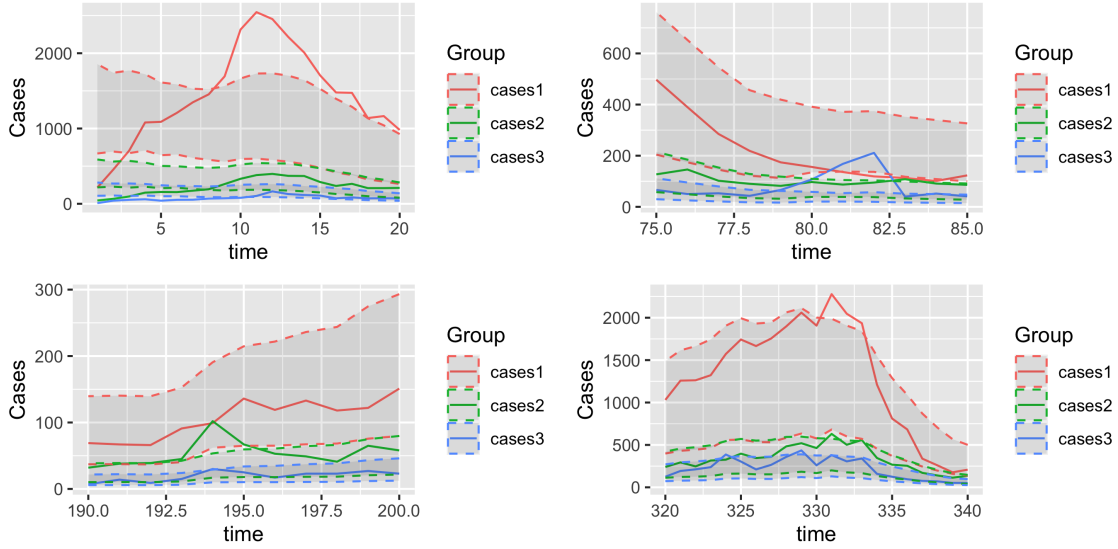
B Model Misspecification

Figure 8: Zoomed-in time-Cases plot at time $t = 0-20, 75-85, 190-200$ and $320-340$. In the visualization, solid lines represent the true reported cases from the original rotavirus dataset, while the shaded strips with dashed boundaries-colored according to different groups—indicate enlarged "confidence intervals" for these cases. Specifically, the upper boundary of each strip corresponds to the upper 97.5% percentile of 36 replicates of the filtered accumulated number of cases, H_{kt}, for $k = 1, 2, 3$, and $t = 1, 2, ..., 416$, as introduced in Section 3.2. This boundary is further adjusted by the upper 97.5% percentile of the reporting rate q_t, estimated by Whitehouse et al. (2023) in Table 3. The reporting rate q_t is assumed to follow a truncated normal distribution with a mean of 0.07 and a standard deviation of $\sigma_q = 0.021$, resulting in a lower 2.5% percentile of 0.029 and an upper 97.5% percentile of 0.111. The methodology for establishing the lower boundary of the strip mirrors that of the upper boundary. Although this approach is not entirely rigorous, each strip is intended to encompass our extended confidence interval for the true reported cases. Consequently, data points that fall outside these strips give us a intuition of where the model misspecifications may occur.

As illustrated in Figure 3, shortfalls are observed at times $t = 1, 2, 3, 11, 81, 194$, and $325-333$. These discrepancies are primarily attributable to outliers that the model fails to accommodate. In the displayed results, it is immediately apparent that at the initial times $t = 1, 2, 3$ all the true reported cases deviated from the expected range, falling outside the designated strip. Specifically, at time $t = 11$, the model is misspecified for case 1, depicted by the solid red line. Similarly, at times $t = 81, 82$, the model is misspecified for case 3, as indicated by the solid blue line that extended beyond the strip's boundaries. At $t = 194$, the solid green line denotes case 2 as where the model is misspecified. During the period from $t = 325 - 333$, all three observed cases are where the model misspecifications happen. These deviations provide a clear illustration of the likelihood shortfall of the sequential Monte Carlo (SMC) method when the model does not align with the data, suggesting a model misspecification.

References

Andrieu, C., Doucet, A., and Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 72(3):269–342.

Avitzour, D. (1995). Stochastic simulation Bayesian approach to multitarget tracking. *IEE Proceedings - Radar, Sonar and Navigation*, 142:41–44(3).

Battineni, G., Chintalapudi, N., and Amenta, F. (2020). SARS-CoV-2 epidemic calculation in Italy by SEIR compartmental models. *Applied Computing and Informatics*, ahead-of-print.

Becker, A. D. and Grenfell, B. T. (2017). tsiR: An R package for time-series susceptible-infected-recovered models of epidemics. *PLOS ONE*, 12(9):1–10.

Biswas, M. H. A., Paiva, L. T., and de Pinho, M. (2014). A SEIR model for control of infectious diseases with constraints. *Mathematical Biosciences Engineering*, 11(4):761–784.

Bjørnstad, O. N., Finkenstädt, B. F., and Grenfell, B. T. (2002). Dynamics of measles epidemics: Estimating scaling of transmission rates using a time series SIR model. *Ecological Monographs*, 72(2):169–184.

Bretó, C., Ionides, E. L., and King, A. A. (2019). Panel data analysis via mechanistic models. *Journal of the American Statistical Association*, 115(531):1178–1188.

CDC (2020). Available from: http://www.cdc.gov/rotavirus/about/index.html.

Chopin, N. and Papaspiliopoulos, O. (2020). *An introduction to Sequential Monte Carlo.* Springer Series in Statistics, 0172-7397. Springer, 1st edition.

Courant, R. (1943). Variational methods for the solution of problems of equilibrium and vibrations. *Bulletin of the American Mathematical Society*, 49(1):1–23.

Doan, T. N., Kong, D. C. M., Marshall, C., Kirkpatrick, C. M. J., and McBryde, E. S. (2015). Characterising the transmission dynamics of Acinetobacter baumannii in intensive care units using hidden Markov models. *PLOS ONE*, 10(7):1–15.

Doucet, A., de Freitas, N., and Gordon, N. (2001). *Sequential Monte Carlo Methods in Practice.* Springer New York, New York, NY.

Doucet, A., Godsill, S., and Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10(3):197–208.

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162.

Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica*, 57(6):1317–39.

Gneiting, T., Balabdaoui, F., and Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 69(2):243–268.

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

Gordon, N., Salmond, D., and Smith, A. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F (Radar and Signal Processing)*, 140:107–113(6).

Hadamard, J. (1908). Mémoire sur le problème d'analyse relatif à l'équilibre des plaques élastiques encastrées. *Mémoires présentés par divers savants éstrangers à l'Académie des Sciences de l'Institut de France*, 33.

He, D., Ionides, E. L., and King, A. A. (2010). Plug-and-play inference for disease dynamics: measles in large and small populations as a case study. *Journal of The Royal Society Interface*, 7(43):271–283.

He, D., Lin, L., Artzy-Randrup, Y., Demirhan, H., Cowling, B. J., and Stone, L. (2023). Resolving the enigma of Iquitos and Manaus: A modeling analysis of multiple COVID-19 epidemic waves in two Amazonian cities. *Proceedings of the National Academy of Sciences*, 120(10):e2211422120.

Houtekamer, P. L. and Mitchell, H. L. (1998). Data assimilation using an ensemble Kalman filter technique. *Monthly Weather Review*, 126(3):796 – 811.

Hulin, W. and Wai-Yuan, T. (2000). Modelling the HIV epidemic: A state-space approach. *Mathematical and Computer Modelling*, 32(1):197–215.

Ionides, E. L., Bhadra, A., Atchadé, Y., and King, A. (2011). Iterated filtering. *The Annals of Statistics*, 39(3):1776 – 1802.

Jiang, S., Maggard, K., Shakeri, H., and Porter, M. D. (2021). An application of the partially observed Markov process in the analysis of transmission dynamics of COVID-19 via wastewater. In *2021 Systems and Information Engineering Design Symposium (SIEDS)*, pages 1–6.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45.

King, A. (2024). Getting Started with pomp: https://kingaa.github.io/pomp/vignettes/getting_started.html#Partially_observed_Markov_process_(POMP)_models.

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.

Kitagawa, G. (1987). Non-Gaussian state-space modeling of nonstationary time series. *Journal of the American Statistical Association*, 82(400):1032–1041.

Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, 5(1):1–25.

Lambert, S. B., Faux, C. E., Hall, L., Birrell, F. A., Peterson, K. V., Selvey, C. E., Sloots, T. P., Nissen, M. D., and Grimwood, K. (2009). Early evidence for direct and indirect effects of the infant rotavirus vaccine program in Queensland. *Med J Aust*, 191(3):157–160.

Lemaréchal, C. (2012). Cauchy and the gradient method. *Doc Math Extra*, pages 251–254.

Liu, J. S. (2008). *Monte Carlo Strategies in Scientific Computing*. Springer Publishing Company, Incorporated.

Michael A. Newton, Peter Guttorp, S. C. R. A. and Abkowitz, J. L. (1995). Stochastic modeling of early Hematopoiesis. *Journal of the American Statistical Association*, 90(432):1146–1155.

Nielsen, S. F. (2000). The stochastic EM algorithm: Estimation and asymptotic results. *Bernoulli*, 6(3):457–489.

Pitt, M. K. and Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, 94(446):590–599.

Rabiner, L. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286.

Rubin, D. B. (1987). The calculation of posterior distributions by data augmentation: Comment: A noniterative Sampling/Importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: The SIR algorithm. *Journal of the American Statistical Association*, 82(398):543–546.

Schmidler, S. C., Liu, J. S., and Brutlag, D. L. (2000). Bayesian segmentation of protein secondary structure. *Journal of Computational Biology*, 7(1-2):233–248.

Smith, A. F. M. and Gelfand, A. E. (1992). Bayesian statistics without tears: A Sampling-Resampling perspective. *The American Statistician*, 46(2):84–88.

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.

Wills, A. G. and Schön, T. B. (2023). Sequential Monte Carlo: A unified review. *Annual Review of Control, Robotics, and Autonomous Systems*, 6(Volume 6, 2023):159–182.