Estimation for Disease Models Across Scales

by

Wilmer O. Martinez Rivera

A Dissertation Presented in Partial Fulfillment
of the Requirement for the Degree
Doctor of Philosophy

Approved July 2022 by the
Graduate Supervisory Committee:

John Fricks, Chair
Mark Reiser
Shuang Zhou
Dan Cheng
Shiwei Lan

ARIZONA STATE UNIVERSITY

August 2022

ABSTRACT

Tracking disease cases is an essential task in public health; however, tracking the number of cases of a disease may be difficult not every infection can be recorded by public health authorities. Notably, this may happen with whole country measles case reports, even such countries with robust registration systems. Eilertson et al. (2019) propose using a state-space model combined with maximum likelihood methods for estimating measles transmission. A Bayesian approach that uses particle Markov Chain Monte Carlo (pMCMC) is proposed to estimate the parameters of the non-linear state-space model developed in Eilertson et al. (2019) and similar previous studies. This dissertation illustrates the performance of this approach by calculating posterior estimates of the model parameters and predictions of the unobserved states in simulations and case studies. Also, Iteration Filtering (IF2) is used as a support method to verify the Bayesian estimation and to inform the selection of prior distributions.

In the second half of the thesis, a birth-death process is proposed to model the unobserved population size of a disease vector. This model studies the effect of a disease vector population size on a second affected population. The second population follows a non-homogenous Poisson process when conditioned on the vector process with a transition rate given by a scaled version of the vector population. The observation model also measures a potential threshold event when the host species population size surpasses a certain level yielding a higher transmission rate. A maximum likelihood procedure is developed for this model, which combines particle filtering with the Minorize-Maximization (MM) algorithm and extends the work of Crawford et al. (2014).

To my wife Ludy and my kids Juana and Sergio.

ACKNOWLEDGEMENTS

Firstly, I would like to express my deepest gratitude to Dr. John Fricks for all the time he has dedicated to guiding and helping me in my progress as a research statistician. Also, I would like to thank my committee members for their support.

I want to thank my wife for her endless support and my kids, who have been my motivation and inspiration. Also, I want to thank my mom, who always encouraged me to follow my dreams. And lastly, I want to thank my older brother, who opened the door for my family to continue our education.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

Chapter 1

INTRODUCTION

People have studied disease modeling from both statistical and mathematical modeling perspectives. This dissertation tries a hybrid approach where we use two statistical methodologies in the context of mathematical models. Both of them use state-space models. The first model is in the context of Susceptible-Infected-Recovery (SIR) models, and the second one is in the context of birth-death models.

When a population is exposed to a pathogen and its potential transmission, a state-space model may approximate the underlying reality based on regular reports (number of cases of infected or deaths). This type of modeling could become pertinent to avoiding high number of cases or fatalities. For instance, models for disease dynamics can be used to predict unobservable quantities and may help us understand how the diseases spread based on different interventions.

Developing effective prevention campaigns usually depends on the understanding of the transmission phenomenon. However, uncovering precise mechanism of transmission is challenging. One way to model disease incidence is by using non-linear state-space models. When the specification of a state-space model mimics the true incidence in the population, it typically demands a complex structure and a significant number of parameters to be estimated. The estimation process may be a difficult task for standard methods. Sequential Monte Carlo, also called Particle Filtering, are simulation-based methods for filtering and smoothing that may facilitate the estimation process in non-linear state-space models. In addition, more complex computational methods such as Particle Markov Chain Monte Carlo (pMCMC) and Iteration Filtering (IF) are methods based on Bayesian framework and likelihood

maximization, respectively.

The outline of this dissertation is as follows. Chapter 2 provides a background of the pMCMC and the iteration filtering (IF2) computational tools to calculate Bayesian and maximum likelihood estimators. This adaptation considers two approaches, one by following the R package *pomp* and a second one by using my implementation. Chapter 3 gives a presentation that uses both the pMCMC and the iteration filtering (IF2) computational tools to approximate the Bayesian and maximum likelihood estimators for the parameters of the model proposed by Eilertson *et al.* (2019) and to compare with their results. This paper explores a non-linear state-space model to estimate measles transmission at the country level. Mainly, this chapter is primarily a presentation of a method to perform Bayesian analysis of the whole country presenting the model of Eilertson *et al.* (2019).

Chapter 4 features a paper that proposes a model to understand the dynamics of the interaction of the population of a disease vector and host population. A birth-death process is used to model the vector population. In addition, the model includes an observational model that is the count of cases in the host population. The host population model incorporates a threshold parameter that denotes an additional infection rate when the vector population size passes above certain level. This model also assumes that the vector population is unobserved, and the available data is the number of reported infection cases from the affected population. Further, this chapter offers a novel estimation procedure based on the combination between particle filters and the MM algorithm which is a generalization of the method presented in Crawford *et al.* (2014). Finally, some simulation studies and an extension of the proposed model that applies the model to cholera deaths reported in Bengal's Dacca district from 1891 to 1940 illustrate the approach. This data set presents a seasonal behavior that is considered in the specification of the birth-death model by means of

a sinusoidal function.

Chapter 2

BACKGROUND

2.1 Introduction

In this chapter, we are going to give some background on the types of models that will be used in Chapter 3, disease dynamics. Non-linear state-space models are a framework for various methods that allow an approximation to the true incidence in a population even when the reports are partially or imperfectly observed. Eilertson *et al.* (2019), based on previous works, propose a state-space model for the estimation of measles transmission, which includes a fairly high number of parameters given the length of the data (Simons *et al.* (2012); Chen *et al.* (2012)). The authors used particle filters to predict the unobserved states of the system, which also gives a Monte Carlo approximation of the likelihood function which is then maximized directly over this Monte Carlo generated function to arrive at parameter estimates.

Alternatively, particle Markov chain Monte Carlo (pMCMC) developed in Andrieu *et al.* (2010) allows for Bayesian estimation of model parameters while exploiting particle filtering. In addition, iteration filtering (IF) methods use particle filters within a maximum likelihood framework (Ionides *et al.* (2006, 2015)). This differs from the Eilertson *et al.* (2019) approach because IF does not need to do explicit maximization. Here, we describe both a Bayesian framework while using pMCMC to sample from posteriors and the iteration filtering (IF2) for comparison. This approach will bypass the well-known downsides of optimizing over a Monte Carlo approximated likelihood function.

This chapter is organized as follows: section 2.2 includes the conceptual back-

ground of the SMC and MCMC algorithms and how they are combined to produce the pMCMC method to enable Bayesian inference. Also, this section includes a description of the iteration filtering algorithm. To familiarize ourselves with these methods, section 2.3 shows a toy example that illustrates the use of the SMC and PMCMC methods. Finally, section 2.4 collects the main conclusions.

2.2 Some Techniques to Predict States and Estimate Parameters in a State Space Model (SSM)

Given the complexity of obtaining analytic solutions in the system of non-linear and non-Gaussian SSM, filtering and smoothing can be challenging to perform. Many algorithms such as Important sampling (IS), Sequential Monte Carlo (SMC), Markov chain Monte Carlo (MCMC), or the combination of some of them approximate solutions for the predicting and the updating steps, see Tsay and Chen (2018). These methods are just a short list of the many procedures there are available for this issue.

2.2.1 SMC

SMC is one of the main tools used for Monte Carlo statistical inference; some of its applications are known in the literature as particle filtering, see Doucet *et al.* (2010). Based on Monte Carlo approximations, SMC accomplishes the prediction and updating steps by focusing on conditional distributions Douc *et al.* (2014). Suppose $\{X_{t-1}^{(1)}, ..., X_{t-1}^{(m)}\}$ are a sample with each of the $X_{t-1}^{(j)}$ follows a density $p_\theta(x_{t-1}|Y_{1:t-1})$ where $Y_{1:t} = \{Y_1, ..., Y_t\}$. At time step t, we want to get samples $\{X_t^{(1)}, ..., X_t^{(m)}\}$ following the distribution $p_\theta(x_t|Y_{1:t})$ where $Y_{1:t}$ includes a new observation Y_t. To perform this task, two steps can be developed. First move into the space of x_t from x_{t-1} by keeping the observations till $t-1$, i.e. $Y_{1:t-1}$, this is called the prediction step. To obtain a sample from $p_\theta(x_t|Y_{1:t-1})$, we use the following

$$
\begin{aligned}
p_\theta(x_t, x_{t-1}|Y_{1:t-1}) &= p_\theta(x_t|x_{t-1}, Y_{1:t-1}) p_\theta(x_{t-1}|Y_{1:t-1}) \\
&= \underbrace{p_\theta(x_t|x_{t-1})}_{\text{“state equation”}} \underbrace{p_\theta(x_{t-1}|Y_{1:t-1})}_{\text{“performed at } t-1\text{”}} .
\end{aligned} \tag{2.1}
$$

Suppose $X_t^{(j)}$ follows the distribution $f_\theta(x_t|X_{t-1}^{(j)})$. Thus, $\{(X_t^{(1)}, X_{t-1}^{(1)}), ..., (X_t^{(m)}, X_{t-1}^{(m)})\}$ are random samples following the distribution $p_\theta(x_t, x_{t-1}|Y_{1:t-1})$. Now to perform the second step, which is drawing samples from $p_\theta(x_t|Y_{1:t-1}, Y_t)$ including the observation Y_t, important sampling can be used. Since the marginal samples $\{X_t^{(1)}, ..., X_t^{(m)}\}$ are generated from the trial distribution $p_\theta(x_t|Y_{1:t-1})$ and the target distribution is $p_\theta(x_t|Y_{1:t})$ then by using a weight function we obtain

$$
\begin{aligned}
w_t(x_t) = \frac{p_\theta(x_t|Y_{1:t-1}, Y_t)}{p_\theta(x_t|Y_{1:t-1})} &\propto \frac{p_\theta(x_t, Y_t|Y_{1:t-1})}{p_\theta(x_t|Y_{1:t-1})} \\
&= \frac{p_\theta(x_t|Y_{1:t-1}) p_\theta(Y_t|X_t, Y_{1:t-1})}{p_\theta(x_t|Y_{1:t-1})} = p_\theta(Y_t|X_t).
\end{aligned} \tag{2.2}
$$

Here we introduce some specific forms for some of the conditional densities $p_\theta(\cdot|\cdot)$ which represents a generic form, that is $p_\theta(y_t|x_t) = g_\theta(y_t|x_t)$, and $p_\theta(x_t|x_{t-1}) = f_\theta(x_t|x_{t-1})$. These forms need to be defined to do the calculations. Thus, we can notice that $w_t(X_t^{(j)}) = g_\theta(Y_t|X_t^{(j)})$ supplies an evaluation on how likely we observe Y_t if the state x_t is truly located at $X_t^{(j)}$. To continue moving from t to $t+1$ we can check that $w_{t+1} = w_t g_\theta(Y_{t+1}|X_{t+1})$, for details see Tsay and Chen (2018), page 404. Hence, the full distribution of the latent process can be predicted recursively starting from a defined initial state samples $\{X_0^{(1)}, ..., X_0^{(m)}\}$ following $\mu_\theta(\cdot)$. Then samples, at $t = 1$, $\{X_1^{(1)}, ..., X_1^{(m)}\}$ are generated where each of the $X_1^{(j)}$ follows a density $f_\theta(x_1|X_0^{(j)})$, and the weights are calculated $w_1(X_1^{(j)}) = g_\theta(Y_1|X_1^{(j)})$.

Now discussing how to estimate the parameters $\theta \in \Theta$ following the SMC techniques, the evaluation of the likelihood function is required to obtain maximum likelihood estimators or Bayesian estimators, thus

$$L_{Y_{1:T}}(\theta) = p_\theta(Y_{1:T}) = \int p_\theta(x_{1:T}, Y_{1:T}) dx_{1:T} \tag{2.3}$$

where $t = 1, ..., T$ with T being the total number of observations, and

$$p_\theta(x_{1:T}, Y_{1:T}) = \underbrace{\mu_\theta(x_0)}_{\text{“initial conditions”}} \prod_{t=1}^{T} \underbrace{f_\theta(x_t|x_{t-1})}_{\text{“time evolution process”}} \prod_{t=1}^{T} \underbrace{g_\theta(Y_t|x_t)}_{\text{“observation process”}} . \tag{2.4}$$

We note that $f_\theta(x_t|x_{t-1})$ defines the state equation or the time evolution process Andrieu *et al.* (2010). Using the fact that:

$$L_{Y_{1:T}}(\theta) = p_\theta(Y_1) \prod_{t=2}^{T} p_\theta(Y_t|Y_{1:t-1}), \tag{2.5}$$

through SMC, we can estimate the likelihood as:

$$\hat{L}_{Y_{1:T}}(\theta) = \prod_{t=1}^{T} \frac{1}{N} \sum_{j=1}^{N} w_t(X_{1:T}^{(j)}), \tag{2.6}$$

where N is the number of particles in the SMC step. Next, we discuss how PMCMC combines SMC and Markov chain Monte Carlo (MCMC) techniques.

2.2.2 PMCMC

To do Bayesian inference when $\theta \in \Theta$ is an unknown parameter, typically we rely on the posterior distribution

$$p(\theta|Y_{1:T}) \propto \int p_\theta(x_{1:T}, Y_{1:T}) dx_{1:T} \quad \underbrace{p(\theta)}_{\text{“prior density”}} , \tag{2.7}$$

which in the context of non-linear non-gaussian models, $p(\theta|Y_{1:T})$ does not usually admit a close form expression, whereby inference is complex in practice. That is why approximations are necessary. Particle Markov chain Monte Carlo (pMCMC) is

a computational tool that combines SMC and MCMC to do Bayesian inference by sampling from the posterior distribution (Andrieu *et al.* (2010)). Particle MCMC is applied to areas such as epidemiology (Rasmussen *et al.* (2011); Endo *et al.* (2019)), finance (Pitt *et al.* (2012); Lux (2021)), and other fields.

The second part of the pMCMC algorithm is Markov chain Monte Carlo (for MCMC, for example, see Robert and Casella (2009)) particularly Metropolis-Hasting (MH). However, other MCMC techniques could be incorporated, such as block sampling, adaptive MCMC, or Gibbs sampling, although the latter demands some extra requirement in the SMC algorithm, see for instance Lindsten *et al.* (2014). In the context of Andrieu *et al.* (2010) Gibbs sampler requires the named *"the conditional SMC update"*. Under these modifications this algorithm is called in Andrieu *et al.* (2010) *"Particle Gibbs sampler"*. For the rest of this report, we will only focus on the algorithm pMCMC, SMC + MH.

A pseudocode of algorithm pMCMC is given as follow:

Step 1: initialization, $i = 0$,

a) Set an initial parameter value $\theta^{(0)}$,

b) Run a SMC to generate samples $\{X_{1:T}\}$, targeting $p_{\theta^{(0)}}(x_{1:T}|Y_{1:T})$, and the approximate partial likelihood $\hat{p}_{\theta^{(0)}}(Y_{1:T})$. Randomly choose one trajectory $X_{1:T}^{(0)}$ from $\{X_{1:T}\}$.

Step 2: for iteration $i \geq 1$,

a) Sample a new parameter $\theta^{*(i)}$ from the proposal distribution $q(\theta^{*(i)}|\theta^{(i-1)})$ based on the previous value $\theta^{(i-1)}$,

b) Run a SMC to generate samples $\{X_{1:T}^{*}\}$, targeting $p_{\theta^{*(i)}}(x_{1:T}|Y_{1:T})$, and the approximate marginal likelihood $\hat{p}_{\theta^{*(i)}}(Y_{1:T})$. Randomly choose one

trajectory $X^{*(i)}_{1:T}$ from $\{X^{*}_{1:T}\}$ as candidate for a MCMC sample,

c) Compare the marginal likelihood with that in the previous iteration $\hat{p}_{\theta^{(i-1)}}(Y_{1:T})$. With probability

$$min\left\{1, \frac{\hat{p}_{\theta^{*(i)}}(Y_{1:T})p(\theta^{*(i)})}{\hat{p}_{\theta^{(i-1)}}(Y_{1:T})p(\theta^{(i-1)})}\frac{q(\theta^{(i-1)}|\theta^{*(i)})}{q(\theta^{*(i)}|\theta^{(i-1)})}\right\} \tag{2.8}$$

updated $\theta^{(i)} = \theta^{*(i)}$, $X^{(i)}_{1:T} = X^{*(i)}_{1:T}$, and $\hat{p}_{\theta^{*(i)}}(Y_{1:T}) = \hat{p}_{\theta^{(i)}}(Y_{1:T})$. Otherwise, keep the values from the previous iteration: $\theta^{(i)} = \theta^{(i-1)}$, $X^{(i)}_{1:T} = X^{(i-1)}_{1:T}$, and $\hat{p}_{\theta^{(i)}}(Y_{1:T}) = \hat{p}_{\theta^{(i-1)}}(Y_{1:T})$.

This pseudo-code completes our discussion of pMCMC. Now, we will discuss an alternative method in the likelihood framework that exploits particle filtering to approximate the likelihood.

2.2.3 Iteration Filtering IF2

Iteration filtering is a method that combines iterations or repetitions of a particle filter with perturbation of model parameters to maximize the likelihood, see Ionides *et al.* (2015). This combination improves the numerical stability of data cloning, a method created by Lele *et al.* (2007), where SMC computations numerically approximate the Bayes map. Thus, the iteration filtering algorithm provides a foundation for stable algorithms, and it is a generalization of the data cloning method. Iteration Filtering is mainly helpful in the context of complex models for which global likelihood optimization is challenging. In such cases, Ionides *et al.* (2015) recommend running multiple searches and continuing each search to obtain narrower limits of available computations.

In the context of (2.4), the pseudocode of IF2 is given as follow:

Step 1: Inputs and initialization

a) Inputs: number the iterations M, and number of particles J.

b) Initialization: $\{X_0^j, j = 1, \ldots, J\}$ following $\mu_\theta(\cdot)$, and $\{\Theta_0^j, j = 1, \ldots, J\}$ following $h(\theta|\sigma_m^2)$, with $m = 1, ..., M$, and perturbation scale $\sigma_{1:M}^2$

Step 2: 1. For m in $1 : M$

2. $\Theta_{0,m}^{F,j} \sim h\left(\theta \mid \Theta_{m-1}^j; \sigma_m^2\right)$ for j in $1 : J$

3. $X_{0,m}^{F,j} \sim \mu\left(x_0; \Theta_{0,m}^{F,j}\right)$ for j in $1 : J$

4. For t in $1 : T$

5. $\Theta_{t,m}^{P,j} \sim h\left(\theta \mid \Theta_{t-1,m}^{F,j}, \sigma_m^2\right)$ for j in $1 : J$

6. $X_{t,m}^{P,j} \sim f\left(x_t \mid X_{t-1,m}^{F,j}; \Theta_{t,m}^{P,j}\right)$ for j in $1 : J$

7. $w_{t,m}^j = g\left(Y_t \mid X_{t,m}^{P,j}; \Theta_{t,m}^{P,j}\right)$ for j in $1 : J$

8. Draw $k^{1:J}$ with $P\left[k^j = i\right] = w_{t,m}^i / \sum_{u=1}^J w_{t,m}^u$

9. $\Theta_{t,m}^{F,j} = \Theta_{t,m}^{P,k^j}$ and $X_{t,m}^{F,j} = X_{t,m}^{P,k^j}$ for j in $1 : J$

10. End For

11. Set $\Theta_m^j = \Theta_{T,m}^{F,j}$ for j in $1 : J$

12. End For,

where M is the number of iterations, J number of particles, $\{\Theta_m^j, j = 1, ..., J\}$ the initial parameter swarm, $h(\theta|\sigma_m^2)$ the perturbation density, $\sigma_{1:M}^2$ the perturbation scale, this is by taking a random walk with smaller variance. This procedure shows that the steps between 4 and 10 represent the particle filter algorithm applied to the extended model where the parameters are stochastic perturbed, for each $t = 1, ..., T$ with T being the total number of observations. The M loop iterates the particle filter with decreasing perturbations of the variance parameter. The superscripts F and P

denote solutions to the filtering and the prediction problem, respectively. The weights $w_{t,m}^j$ denotes the likelihood of the data at time t for particle j in filtering iteration m. Thus, the output is the final parameter swarm, $\{\Theta_M^j, j = 1, ..., J\}$. This final parameter swarm is recycled by taking the mean, for instance, for each parameter, respectively which constitute the starting parameters for the next iteration. For this notation, we eliminate the parenthesis from the superscripts j for simplicity. Now, let us look at a simulation example to illustrate SMC and PMCMC.

2.3 Simulation Example

To illustrate how SMC works and how PMCMC exploits the strengths of SMC, we developed a toy simulation. The simulation example is defined as follows

$$\begin{aligned} F_\theta(X_{t-1}) &= aX_{t-1} + bX_{t-1}/(1 + X_{t-1}^2)) + \gamma, \\ X_t &= F_\theta(X_{t-1}) + W_t. \end{aligned} \tag{2.9}$$

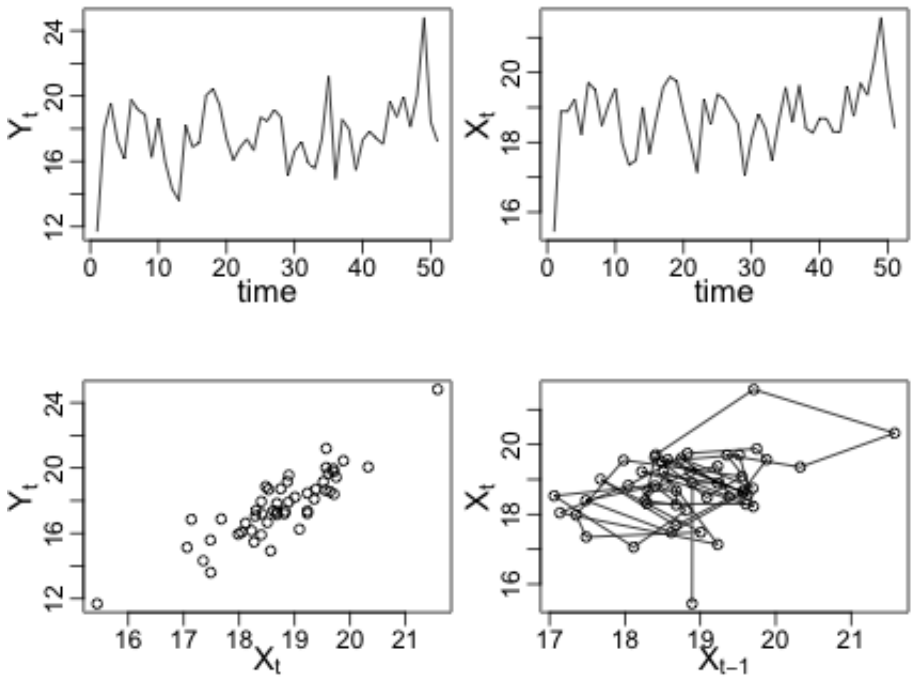
$$\begin{aligned} H(X_t) &= X_t^2/20, \\ Y_t &= H(X_t) + V_t, \end{aligned} \tag{2.10}$$

where $X_0 \sim N(2\gamma, 1)$, with $W_t \sim N(0, 1)$ independent of $V_t \sim N(0, 1)$ and each sequence independent of X_0. And, $\theta = (a = .5, b = 25, \gamma = 8)$ [1] . Figure 4.1 shows an output of this simulation with a sample size of $T = 50$. Now, from this simulation data, we are going to follow both the SMC and MCMC algorithms as we mentioned in section 2.2. To get results by using these two techniques in R there are several packages that allow to implement those such as *LibBi* (Murray (2013)), *Nimble* (de Valpine *et al.* (2017)), *pomp* (King *et al.* (2016)) , *RBi* which is an interface to LibBi,

[1] Taken from Douc *et al.* (2017), page 297.

among others. Particularly, here, we use the package *pomp* (King *et al.* (2016)) for both pMCMC and IF2, and we developed our own code to contrast our results and to verify our own implementation.

Figure 2.1: Simulation of X_t and Y_t for $t = 1, ..., 200$



2.3.1 Particle Filter and Comparison

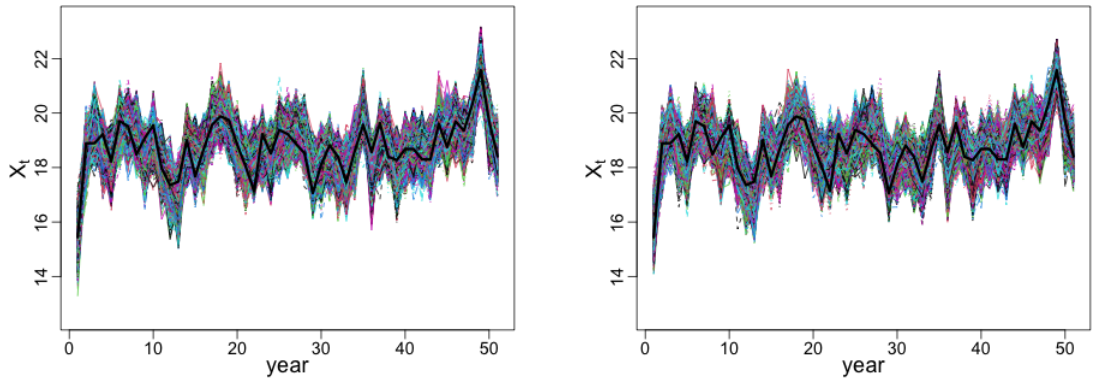
For the implementation of the SMC or the particle filter with bootstrap resampling according to equations (2.4), (2.9), (2.10) we follow this scheme:

Since the goal of using the SMC algorithm, in the pMCMC method, is to produce a prediction of the hidden process, Figures 2.2 and 2.3 show this prediction for the particle filters (coloring lines) and the true state (black line). To obtain these predictions, we used both our implementation code and the *pomp* R package. We can draw

x_0	initial condition	$x_0 \sim N(2\gamma, 1)$
$f_\theta(x_t \mid x_{t-1})$	is defined as	$x_t \sim N(F(x_{t-1}), 1)$
$g_\theta(y_t \mid x_t)$	is defined as	$y_t \sim N(H(x_t), 1)$
data set $y_{1:50}$		
$\theta = (a, b, \gamma)$		= c(0.5,25,8)
N_p	# particles	= 5000 and 10000

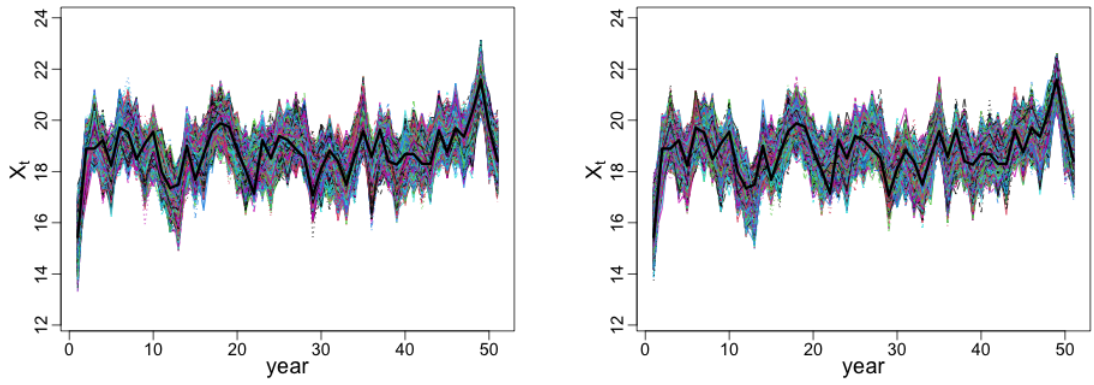
two important conclusions from these results: first, the particle filter covered the true states under the 5000 and 10000 particles cases, and second, our implementation displays results close to *pomp*. Further, the running time is practically the same. In addition, in Figure 2.4, we show the effective sample size for both cases 5000 and 10000 particles, and the results from our implementation look similar to the results from *pomp*. These parallel outcomes allow us to be confident about our performance and our results.

Figure 2.2: Estimation (filter) of the State X_t with 5000 particles colored lines, the black line represents the true State. (a) shows the results by using our own code, and (b) shows the results by using the pomp R package



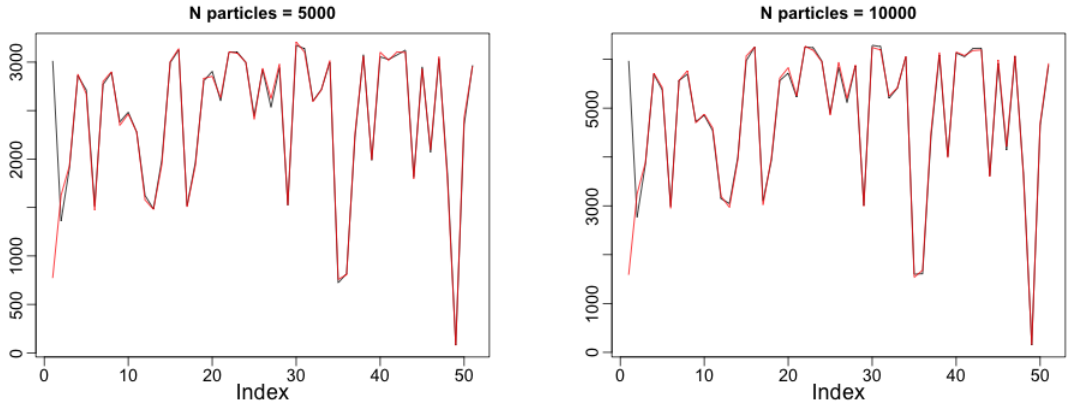
(a) Our (b) Pomp

Figure 2.3: Estimation (filter) of the State X_t with 10000 particles colored lines, the black line represents the true State. (a) shows the results by using our own code, and (b) shows the results by using the pomp R package



(a) Our (b) Pomp

Figure 2.4: Effective sample size (efss) from our implementation (black line), and from pomp (red line)



(a) (b)

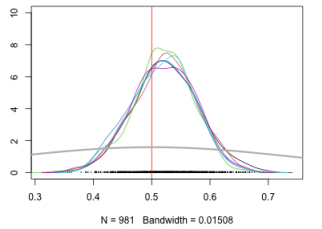
2.3.2 PMCMC and Comparisons

Now, from the previous results and by following the pseudocode given in 2.2.2, we estimate the parameters of the model in (2.9) by combining the SMC and the MCMC algorithms. For the first part of these outcomes we compare the parameter estimation using as a proposal distribution, $q(\theta^{*(i)}|\theta^{(i-1)})$, a normal random walk with diagonal covariance matrix; and independent *Normal prior* distributions with mean $\{0.5, 25, 8\}$ and standard deviation $\{0.52, 12.5, 4\}$ respectively for the parameters $\{a, b, \gamma\}$. In order to check stationarity condition it is recommended to run several Markov chains see Robert and Casella (2009) or Dahlin and Schön (2019). Therefore, by using pomp, we simulated five independent chains with 10^4 particles and 10^5 iterations. We burned in using the first half, and we used a thinning parameter equal to 50. From our implemented procedure, we run one chain. Figure 2.5 shows the main results from the pMCMC algorithm. In figures 2.6a, 2.6b, and 2.6c we present the estimated marginal posterior for a, b, and γ, five coloring lines from pomp and one black line for our implementation. The gray line represents the prior distribution. From these figures, we cannot see appreciably significant differences between *pomp* and ours. Also, the estimated marginal posterior covers the true parameter value (red line). Something to remark on here, regarding the parameter b, is that a transformation like e^b could help to improve the estimation in the sense that we should not restrict the range of the parameter b. Figures 2.6d-2.6f include the trace of the Markov chain. All of them exhibit a behavior approximately stationary, and figures 2.6g-2.6i show the ACF for each Markov chain.

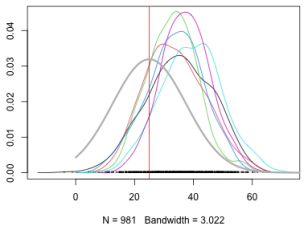
In addition, we obtain results by using an independent *uniform priors* with limits $\{0.05, 2.5, 0.8\}$ and $\{2.5, 125, 40\}$ respectively for the parameters $\{a, b, \gamma\}$. In Figure 2.6 we show these outcomes. We notice that the results in this case are similar to those

we obtained under the normal prior assumption. We also notice that the estimation of the marginal posterior for the parameter b is more precise under the normal prior distribution.

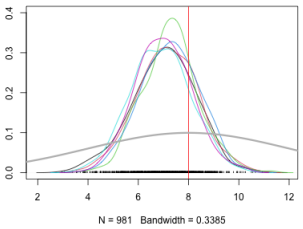
Figure 2.5: Density plots (Figures a, b, and c): Normal prior distribution (gray line), posterior estimations (coloring line by using pomp), and by using our implementation (black line), red line correspond to the true parameter value. Trace plots of the Markov chain (Figures d, e, and f) for the parameters a, b, and γ parameters, respectively. ACF plots for each Markov chain (Figures g, h, and i) for the parameters a, b, and γ parameters, respectively



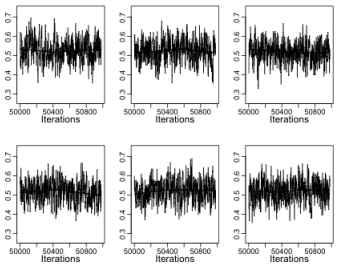
(a) a



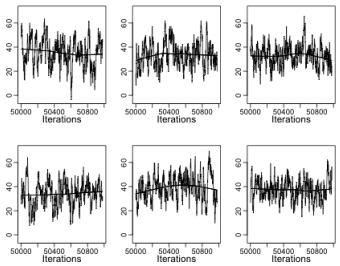
(b) b



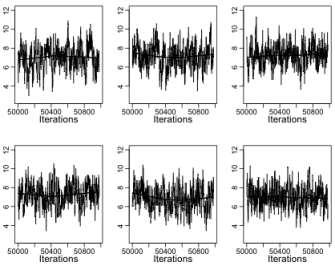
(c) γ



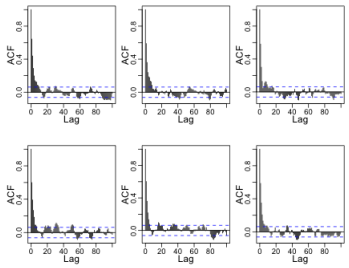
(d)



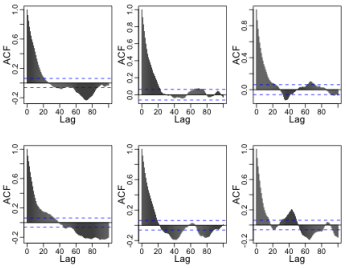
(e)



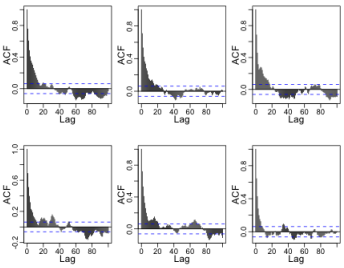
(f)



(g)

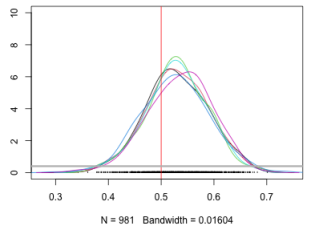


(h)

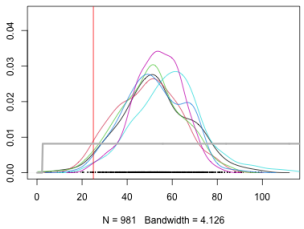


(i)

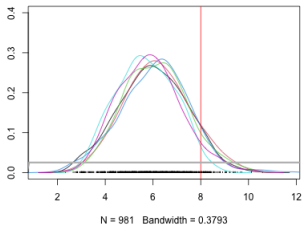
Figure 2.6: Density plots (Figures a, b, and c): Uniform prior distribution (gray line), posterior estimations (coloring line by using pomp), and by using our implementation (black line), red line correspond to the true parameter value. Trace plots of the Markov chain (Figures d, e, and f) for the parameters a, b, and γ parameters, respectively. ACF plots for each Markov chain (Figures g, h, and i) for the parameters a, b, and γ parameters, respectively



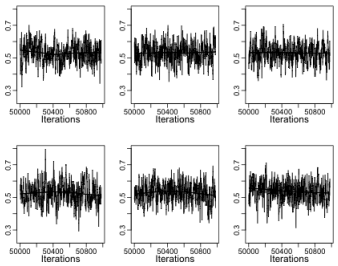
(a) a



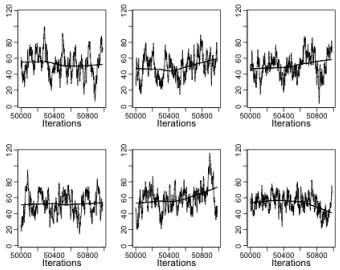
(b) b



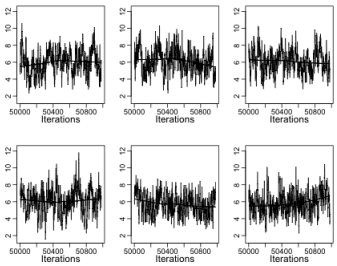
(c) γ



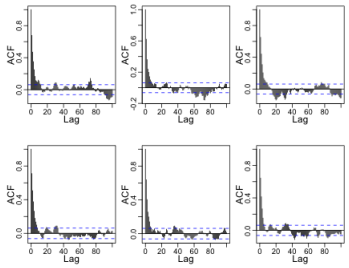
(d)



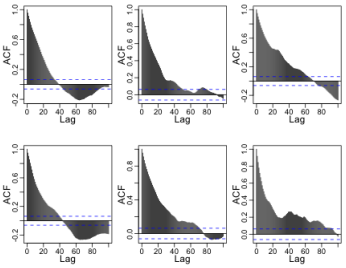
(e)



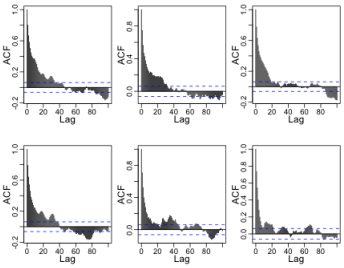
(f)



(g)



(h)



(i)

2.4 Conclusions

By using a simulated example, we evaluate our implementation of the pMCMC method. It was compared with the results obtained from the R package *pomp*. The running time for the SMC or particle filter is similar to time by using *pomp*. Still, the running time for the pMCMC algorithm with our implementation is almost three times the total time by using *pomp*.

Chapter 3

BAYESIAN ESTIMATION OF PARAMETERS FROM WHOLE COUNTRY ANNUAL MEASLES CASES USING PARTICLE MCMC (PMCMC)

3.1 Introduction

Tracking disease cases is an important task in public health, so to do this tracking, there are basic SIR models, see for instance Anderson and May (1991), Bjørnstad *et al.* (2002), or Allen (2008). It is often hard to keep track of the number of cases since not every infection is reported to the public health authorities. That is a problem for public health officials, so we will come up with a method that treats the cases as hidden unobservable quantities. Then, we will have an observation model and put all of these aforementioned quantities within a state space model.

The under-reported surveillance data produces a partially observed time series. Due to the existence of imperfect observations, health professionals often only partially observe the actual cases in a population. Notably, this happens with the measles case reports in whole countries, even those with solid registration systems (Simons *et al.* (2012)). Under these conditions, estimating the true number of disease cases is difficult and problematic for policymakers when planning and setting funding priorities (WHO (2007)).

Linear and non-linear state-space models are a class of models for which standard procedures exists in order to allow a quantification of the distribution of the true number of cases in a population even when the reports are only partially observed. This is known as *Filtering*. State space models combine a state equation, X, and an observation equation Y, where Y is a transformed version of X typically with noise

added. Thus, these models help predict the unobserved number of cases of diseases in a population and the dynamics of disease infection more generally (see, Eilertson *et al.* (2019)). Simons *et al.* (2012) propose a state space model that incorporates surveillance data to estimate incidence and age distribution based on the reported number of measles cases. By using the annual reports from 192 countries, 65 with vital registration and 128 countries with inadequate vital registration data, they concluded that underreporting is estimated to be below 10%. Chen *et al.* (2012), previously produced a methodology for a similar kind of data setup by developing state space methods to predict unobserved measles burden. They were estimating the parameters of the model using an extended Kalman Filter to calculate the likelihood function then maximizing.

Eilertson *et al.* (2019) propose using a state-space model combined with maximum likelihood methods for estimating measles transmission, which includes a relatively high number of parameters given the length of the data. The authors used particle filters to predict the system's unobserved states in particular the number of susceptibles and the number of infecteds, which also produces a Monte Carlo approximation of the likelihood function, that is then maximized directly to arrive at parameter estimates. In addition, they include age classes within the states. Based on this model, without considering the age classes, we propose an alternative Bayesian approach that uses particle Markov chain Monte Carlo (pMCMC) from Andrieu *et al.* (2010) to draw from the posteriors and iteration filtering (IF2) method from Ionides *et al.* (2015). We are going to be using IF2 as a check on the Bayesian approach and also to inform priors. This approach differs from the Eilertson *et al.* (2019) because IF2 has stronger theoretical underpinnings; moreover, it does not need to do explicit maximization. This approach will bypass the well-known downsides of optimizing over a Monte Carlo approximated likelihood function.

This chapter is organized as follows: section 3.2 describes the model proposed by Eilertson *et al.* (2019) removing the age-classes component. Section 3.3 develops a simulation study to explore Bayesian analysis using pMCMC and IF2 as check. Section 3.4 presents a study case implementing the model proposed by Eilertson *et al.* (2019). Finally, section 3.5 states the main conclusions and some recommendations.

3.2 Model Description

The model considered here, from Eilertson *et al.* (2019), is a phenomenological model with some mechanistic aspects for dynamics at the country level to provide a national estimates of the measles disease burden. This model mimics the behavior captured by the conventional SIR-type model. Although the model in Eilertson *et al.* (2019) considers the age effect, we emphasized the application of the simpler version without including the age classes more similar to Chen *et al.* (2012). We evaluate the model via simulation using the covariates that consist of population size, number of births and vaccination rate of the national data of the country of India, and then we do case studies with selected countries. To clarify this model, we write it in a different way than the model is written in Eilertson *et al.* (2019). We use a set of independent uniform random variables and indicator functions to represent the model, as we can see in equations (3.1-3.5). This model is on an annual timescale, where S_t is the number of susceptibles, the people in a population vulnerable to becoming sick at year t, which is a function of the number of susceptibles and the number of infected individuals in the previous year S_{t-1}, and I_{t-1}, respectively. Also, S_t is function of the number of births at year t, B_t, modified by the fraction of immunized people through vaccination V_t. According to Chen *et al.* (2012), V_t is defined as the efficacy of the measles vaccine, which is a linear combination of $V_{1,t}$ and $V_{2,t}$ the vaccine coverage with one and two doses at time t, respectively. Thus, V_t is given as in (3.2).

$$
\begin{aligned}
S_t &= S_{t-1} + B_t(1 - V_t) - I_{t-1}, \\
S_0 &= M \times N_1, \\
M &\sim U(0.01, 0.1),
\end{aligned}
\tag{3.1}
$$

$$
V_t = 0.85V_{1,t}(1 - V_{2,t}) - 0.99V_{1,t}V_{2,t}, \tag{3.2}
$$

For the initial condition S_0, we draw from a uniform random variable, M, between 1% and 10% of the total population, N_1, at $t = 1$ in which the latter being a proxy for the population at $t = 0$. The expected number of infected persons in Eilertson *et al.* (2019) is assumed to be an increasing function of the fraction of the population that is susceptible $\frac{S_t}{N_t}$; therefore, the number of infecteds, I_t, each year is modeled as a binomial draw with parameters S_t and dynamic probability π_t as described in (3.4). In this sense, the form of π_t reflects the indirect protection of herd immunity as a fraction of the population. An illustration of the attack rate π_t as function of the parameters β_0 and β_1 and the fraction S_t/N_t (values between 0 and 1) is provided in Figure 3.2a. This illustration ignores the noise term e_t. This figure allows us to see how much π_t changes as we change the parameters β_0 and β_1. The e_t are a sequence of independent normal random variables, through time, with zero mean and fixed variance σ_e^2. Thus,

$$
I_t = \sum_{j=1}^{S_t} \mathbb{1}_{\pi_t}(j, t), \tag{3.3}
$$

where

$$
\mathbb{1}_{\pi_t}(j, t) = \begin{cases} 1 & \text{if } U_{j,t} < \pi_t \\ 0 & \text{otherwise,} \end{cases}
$$

$$\begin{aligned} \pi_t &= logit^{-1}\left(\beta_0 + \beta_1 \frac{S_t}{N_t} + e_t\right), \\ e_t &\sim N(0, \sigma_e^2), \end{aligned} \tag{3.4}$$

The number of reported cases, C_t, comes from the surveillance data, which we assume as a subset of the actual number of infected I_t because of the effect of under reporting. Thus, C_t follows a conditionally binomial distribution with parameters I_t and probability ρ_r, which is distinct for each country.

$$C_t = \sum_{i=1}^{I_t} \mathbb{1}_{\rho_r}(i,t), \tag{3.5}$$

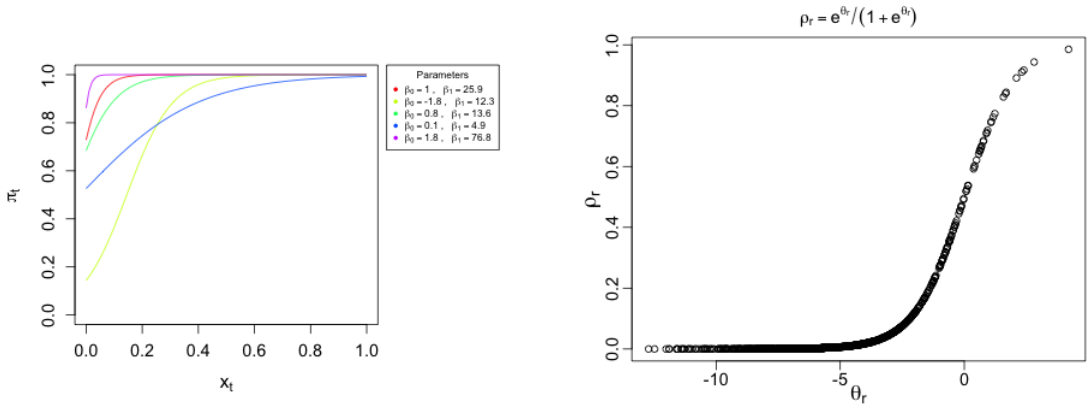
$$\mathbb{1}_{\rho_r}(i,t) = \begin{cases} 1 & \text{if } R_{i,t} < \rho_r \\ 0 & \text{otherwise.} \end{cases}$$

The random variable, M, and the sequences of random variables $U_{j,t}$, e_t, and $R_{i,t}$ are independent from each other and through t, i, and j. In Table 3.1, we summarize all the terms in the model. In the coming sections, we will evaluate the estimators associated with this model. The evaluation will be considered in two simulation cases: first by estimating the three parameters β_0, β_1, and θ_r fixing σ_e^2; and the second by including the fourth parameter σ_e^2. Subsequently, we will use some countries' data sets to evaluate estimations in a real context.

Table 3.1: Description of Model Terms

Term	Description	Type
S_t	Number of Susceptible at time t	Unobserved
I_t	Number of Infected at time t	
B_t	Births at time t	Given
V_t	Fraction of immunized people through vaccination	
N_t	Population size at time t	
C_t	Observed number of cases reported in year t	
π_t	Probability of attack rate in year t	Parameter
β_0	Parameter in attack rate function	
β_1	Parameter in attack rate function	
σ_e^2	Attack rate variation (of e_t)	
ρ_r	Probability of reporting a case	
e_t	Effect of unspecified influences on attack rate	Random variable
M, $U_{j,t}$, and $R_{i,t}$	Independent uniform random variables.	

Figure 3.1: Parameters behavior. (a) shows the curves obtained from equation (3.4) for π_t where x_t represents the fraction of susceptible in a population and by selecting some values for β_0 and β_1. (b) Logistic transformation for parameter ρ_r.



(a) (b)

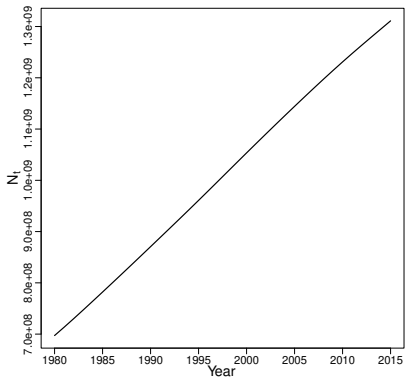
3.3 Simulation Study

In this section, we develop Bayesian statistical methodology for the model presented in the previous section. The Bayesian estimation procedure uses particle MCMC (pMCMC) to generate approximate samples from the posterior. Particle MCMC is a computational tool that enables Bayesian analysis, and IF2 that calculates maximum likelihood estimators allows us to check the fidelity of this analysis, especially in the case of real data, and also it allows us to empirically select priors. In this section, we evaluate the estimation for this model via simulation.

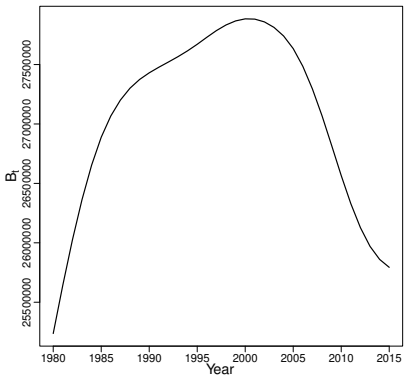
We simulate 36 observations (mimicking the sample size of the real measles data which we will analyze in section 3.4) for $X_t = \left(S_t, I_t\right)$ which is the system equation and therefore unobserved, and $Y_t = C_t$, which is the observed number of cases. Also, we use the covariates B_t, V_t, and N_t (see Table 3.1 and Figures 3.2) from the country of India. For this simulation study we set the values of the vector of parameters as $\Theta = (\beta_0, \beta_1, \theta_r, \sigma_e^2) = (1.3, 30, -5, 0.5^2)$, these values correspond to results obtained in Eilertson *et al.* (2019). To avoid difficulties in the estimation of the parameter representing the probability of reporting a case, ρ_r, since it has a restricted range (0, 1), we used a logistic transformation $\rho_r = e^{\theta_r}/(1 + e^{\theta_r})$. A realization with these settings is displayed in Figure 3.2b.

To perform the estimation we consider two cases: first, we estimate the three parameters β_0, β_1, and θ_r by fixing the parameter σ_e^2 as it is suggested in Eilertson *et al.* (2019), and second we estimate the full parameter vector Θ. Now we consider the first case.

Figure 3.2: Covariates: Populations size (N_t), Number of births (B_t), vaccine rate (V_{1t}), and number of Cases (C_t) . These data are from country of India

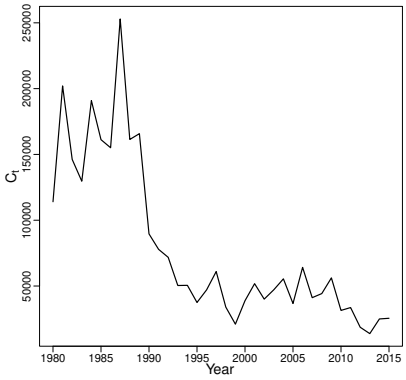


(a) Population size



(b) Number of births

(c) Effective vaccination rate

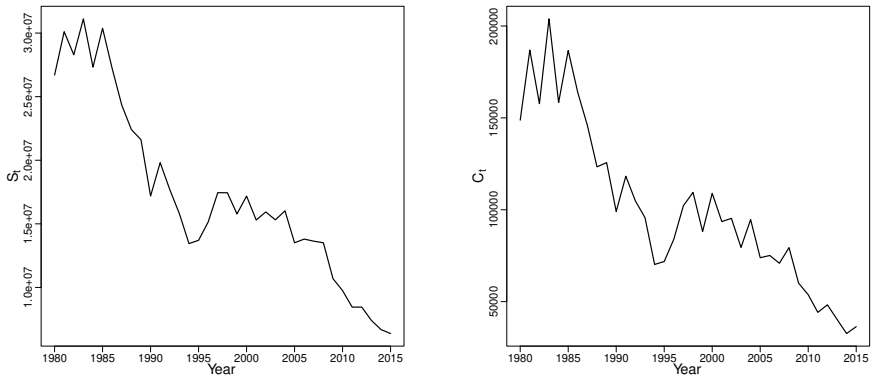


(d) Number of cases

3.3.1 Three Parameters Estimation

A key part of the model is the attack rate in year t, π_t, which includes the parameters β_0, β_1, and σ_e^2. To understand the relationship between these parameters and the attack rate, we create curves of π_t vs. x_t which represent the fraction of susceptibles in the population whose values are between 0 and 1, and with a collection of selected values for β_0 and β_1 without taking into account the effect of e_t, see Figure 3.2a. Also, to understand the behavior of the parameter ρ_r under the logistic transformation, we display a curve in Figure 3.2b. These plots allow us to define values for the three

Figure 3.3: Simulated susceptible population size and number of cases using covariates N_t, B_t, and V_t from the country of India.



(a) Susceptible population (b) Number of cases

parameters to be considered in the current simulation study. Particularly, values for θ_r around -5 provides values for ρ_r around 0.1, and the colored lines in Figure 3.2a provides us some guesses for β_0, β_1.

The iteration filtering (IF2) algorithm, described in section 2.2.3, yields maximum likelihood estimators of the parameters of a partially-observed Markov process. The IF2 performs a specified number of particle-filter iterations. At each iteration, the particle filter is performed on an augmented version of the hidden states which include the parameter vector Θ_t; the algorithm is presented in Ionides *et al.* (2015). In the system equation these parameter values are undergoing a Gaussian random walk with a normal transition density with zero mean and variance following a perturbation scale as we described in the Chapter 2, Section 2.2.3.

In practice, to better explore the parameter space, Ionides *et al.* (2015) recommends using several filtering iterations. The implementation of the IF2 algorithm is carried out by using the R package called *pomp* with the following specifications: we use 20 filtering iterations, 20000 particles, and 450 iterations. Each filtering starts at

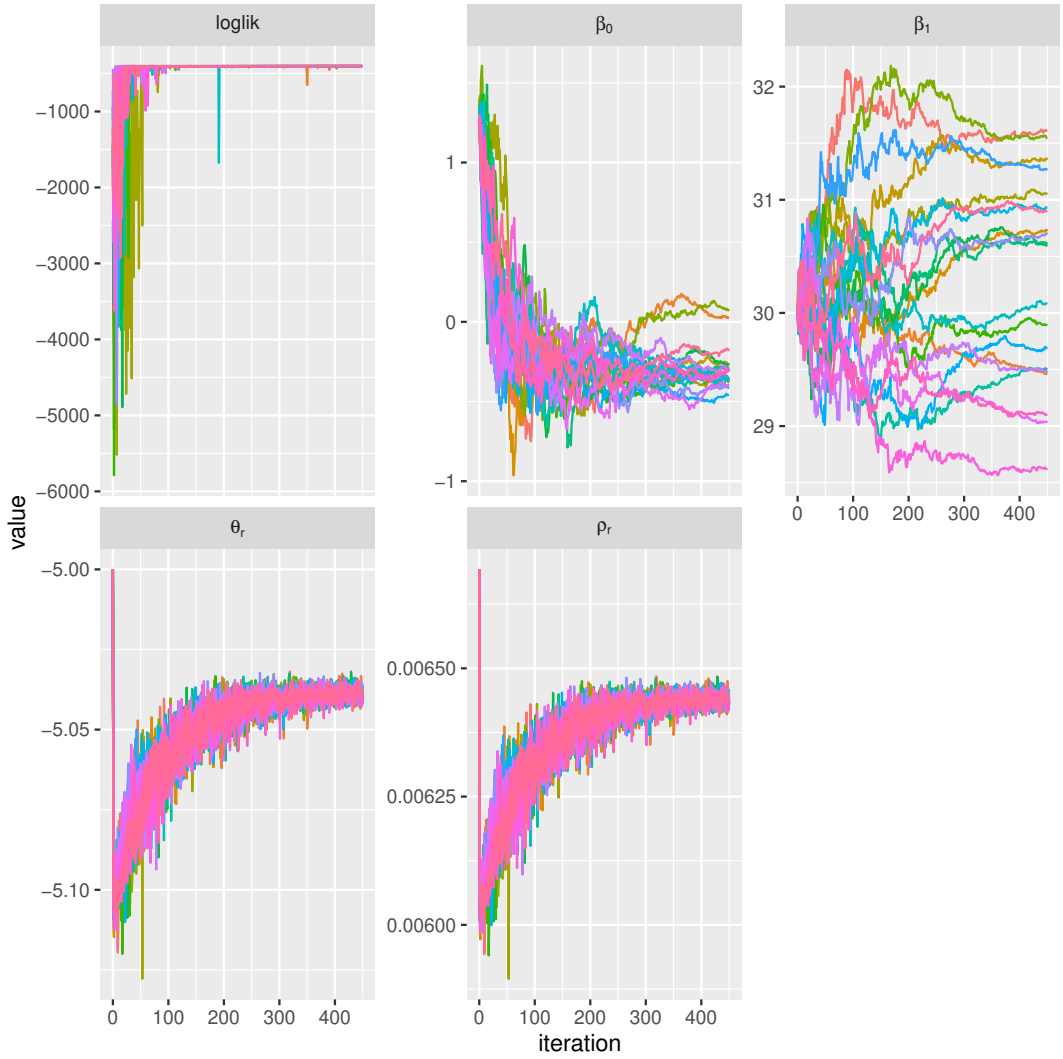
the true parameter values and a cooling parameter of 0.7, which defines the perturbation scale and, in steps of the context of the R package *pomp*, this cooling parameter means that every 70 iterations, the standard deviation of the random walk corresponding to each parameter decreases by half of the current value. The results of the estimation from this method are shown in Figure 3.4. These figures include estimates of the log-likelihood for each filtering iteration (colored lines), the mean of each filter iteration for the three parameters, and the transformation to the original parameter ρ_r.

The pMCMC implemented here, described in section 2.2.2, is the one that uses particle Metropolis-Hastings from Andrieu *et al.* (2010). Particle MCMC combines a particle filter and Metropolis Hasting, a more complete discussion of Metropolis Hasting is carried out in Robert and Casella (2009). First, the particle filter approximates the hidden states (finds the filtering distribution) and obtains estimates of the likelihood. Second, Metropolis-Hastings (MH) generates a Markov chain based on the likelihood estimators provided by the particle filter whose draws are approximately the target distribution, see Dahlin and Schön (2019).

To carry out a Bayesian analysis, we attach prior distributions to our parameters β_0, β_1, and θ_r, the selected priors are distributed independently normal. We choose four sets of hyper-parameters to check if the posterior densities depend substantially on the prior choice. The hyper-parameters selected are: prior 1 centered on the true parameter, prior 2 centered on the true parameter value with double the standard deviation compared to prior 1, prior 3 modified the mean of the β_1 displacing to the left of the true value, and prior 4, all the priors move the mean to the right of the true parameter values, respectively, see Table 3.2. We focus on only these four choices for a running time, but for future work, more choices will be explored.

The choice of the proposals is influential on the performance of the MH algo-

Figure 3.4: Results from IF2 with initial values: β_0 =1.3, β_1 =30, and θ_r =-5



rithm. In practice, proposals with scaling calibration are a set of proposals discussed in Section 6.4 of Robert and Casella (2009). Running the MH with random walk proposals with different scale parameters is good practice to improve the mixing rate of the chains; see example 6.5 in Robert and Casella (2009). Following this recommendation, here we select proposal distribution, $q(\Theta^{*(i)}|\Theta^{(i-1)})$, of a multivariate normal random walk with diagonal covariance matrix $\Sigma = diag(\epsilon_j^2)$, where ϵ_j^2 denotes a vector with three variances and $j = 1, .., 5$, thus we consider five proposals as we summarize

in Table 3.2.

Now by combining the four selected priors and the five selected proposals, we display in Figure 3.5 kernel density estimates from the posterior. Each row, in Figure 3.5, corresponds to a different set of hyper-parameters, and each column represents the parameter which is being estimated. The entries in every row and every column represents the posteriors estimates (colored lines, which are related to the selected proposals, see Table 3.2). Gray lines denote prior densities, and the red line represents the true parameter value. From these results, we can see that the posterior densities do not change substantially across choices of priors, which allows us to conclude that the Bayesian estimation is robust with respect to these priors. In these results, we also include the posterior densities for the original parameter ρ_r and the corresponding prior as follows.

Prior transformation for ρ_r and θ_r. $\theta_r \sim N(\mu_{\theta_r}, \sigma^2_{\theta_r})$.

$$\begin{aligned} \rho_r &= \frac{e^{\theta_r}}{1+e^{\theta_r}}, \\ \theta_r &= log\left(\frac{\rho_r}{1-\rho_r}\right). \\ f_{\rho_r}(\rho_r) &= f_{\theta_r}\left(log\left(\frac{\rho_r}{1-\rho_r}\right)\right)\left|\frac{d}{d\rho_r} log\left(\frac{\rho_r}{1-\rho_r}\right)\right| \\ f_{\rho_r}(\rho_r) &= f_{\theta_r}\left(log\left(\frac{\rho_r}{1-\rho_r}\right)\right)\left(\frac{1}{\rho_r(1-\rho_r)}\right), \end{aligned} \tag{3.6}$$

where $f_{\theta_r}(\cdot)$ is $N(\mu_{\theta_r}, \sigma^2_{\theta_r})$, and $\frac{e^{\theta_r}}{1+e^{\theta_r}}$ is monotonically increasing.

The estimation of the posterior densities is based on the simulation of independent Markov chains with 50000 particles and 10^5 iterations. Based on the trace (the states of the Markov chain) and the estimated ACF of the Markov chain, we burn the 30 thousand initial values and choose a thinning parameter equal to 20, Figures 3.6 - 3.9. From the posterior marginal densities (colored curves), we have that the posterior density covers the true parameter (red line) in most cases. For a few cases, it is not

Table 3.2: Specifications of the Prior and Proposal Distributions

			β_0	β_1	θ_r
Prior 1	Normal	μ	1.3	30	-5
	(grey line)	σ^2	1^2	10^2	0.05^2
Prior 2	Normal	μ	1.3	30	-5
	(grey line)	σ^2	2^2	20^2	0.1^2
Prior 3	Normal	μ	1.3	0	-5
	(grey line)	σ^2	2^2	20^2	0.1^2
Prior 4	Normal	μ	3	46	-4.9
	(grey line)	σ^2	3^2	30^2	0.15^2

		Standard deviation (ϵ_j) proposal distribution		
		β_0	β_1	θ_r
Posterior				
—	$j = 1$	0.05	0.5	0.005
—	$j = 2$	0.1	1	0.01
—	$j = 3$	0.5	5	0.05
—	$j = 4$	1	10	0.1
—	$j = 5$	1.5	15	0.15

covered; however, it is visibly quite close especially given the small sample size (36 sample time points). In addition, Figures 3.10 show the histograms, the scatter plots and the contour lines for the posterior densities estimated under the proposal 3, see Table 3.2, according to the criterion of nearly uncorrelated defined below. From these plots, we can identify a strong correlation between the posterior estimate densities for the parameters β_0 and β_1 for each of the simulated chains.

Now we evaluate the predictive properties of the model and methods. Given the system equation and the observation process in (2.4), the one-step ahead prediction density is given by

$$p_\theta(Y_{t+1}|Y_{1:t}) = \int\int p_\theta(Y_{t+1}|x_{t+1})p_\theta(x_{t+1}|x_t)p_\theta(x_t|Y_{1:t})dx_{t+1}dx_t. \tag{3.7}$$

To produce an estimate of this density, we proceed as follows

1. x_t is approximate by $p_\theta(x_t|Y_{1:t})$, where $j = 1, ..., J$ and J is the total number of particles.

2. $x^{(j)}_{t+1|t}$ is approximate by $f_\theta(x_{t+1}|x_t)$, where $f(\cdot)$ as we specified in Chapter 2.

3. $Y^{(j)}_{t+1|t}$ is approximate by $g_\theta(Y_{t+1}|x_{t+1})$, where $g(\cdot)$ as we specified in Chapter 2.

Thus, at $t+1$ we obtain a cloud of potential values $\{Y_{t+1|t}^{(1)}, Y_{t+1|t}^{(2)}, ..., Y_{t+1|t}^{(J)}\}$. Now, we propose three methods to calculate one-step ahead predictions.

1. Hold θ constant from the IF2 estimates, whose results are called *MLETRAD*.

2. Hold θ constant by taking the MAP estimators, i.e., taking the mode of the posterior estimates, whose results are called *MAPTRAD*.

3. We draw samples θ_i from the posterior estimates, whose results are called *POSTPRED*.

Predictions from methods *MLETRAD* and *MAPTRAD* are performed as follows: We calculate $\{Y_{t+1|t}^{(1)}, Y_{t+1|t}^{(2)}, ..., Y_{t+1|t}^{(J)}\}$, then we take the average of these predictions as a point prediction, this is

$$\hat{Y}_{t+1} = \frac{1}{J}\sum_{j=1}^{J} Y_{t+1|t}^{(j)}. \tag{3.8}$$

With the method *POSTPRED*, we draw sequentially samples θ_i, $i = 1, ..., M$, from the posterior estimates and with each sample we calculate $\{Y_{t+1|t}^{(i,1)}, Y_{t+1|t}^{(i,2)}, ..., Y_{t+1|t}^{(i,J)}\}$ then averaging values across the samples leads to a Monte Carlo approximation to the one-step ahead prediction density, for example see Prado and West (2010) and Leung *et al.* (2019). Thus, we obtain a point prediction $\hat{Y}_{t+1}$.

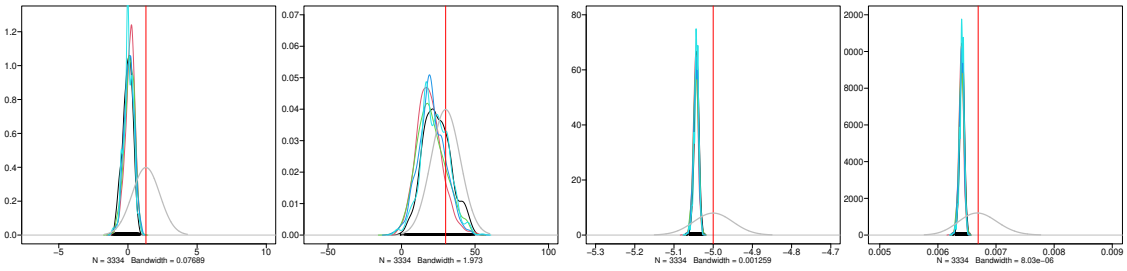
Since we combined different priors and proposals, we obtained a posterior estimate for each combination. To evaluate the predictive properties of the model and methods, we choose one of these posterior estimates. We choose posterior estimates, whose corresponding estimated ACF of the Markov chain exhibits nearly uncorrelated behavior (the posterior densities estimated under proposal 3). In Figures 3.11 we show the results of the one-step ahead predictions. For methods *MLETRAD* and

MAPTRAD, we run the particle filters with 1000 particles and for method *POSTPRED* we run the particle filters with 1000 particles per sample θ_i, with $i = 1, ..., 50$ (see Figures 3.11). The actual number of cases is represented with a red line. Also, in Figures 3.11 we show the graphs for the four evaluated priors. From these results, we conclude that the performance of the one-step ahead predictions under the different priors is similar. For all the cases, the true value is covered by the clouds of particles. To ease the reading of these results, we only include the estimates of a selected number of cases for some time points. In addition, Figures 3.12 include the residuals and studentized residuals. From these plots, we can see a good behavior of the residuals since they fluctuate around zero with a bigger spread of the residuals with the method *POSTPRED*.

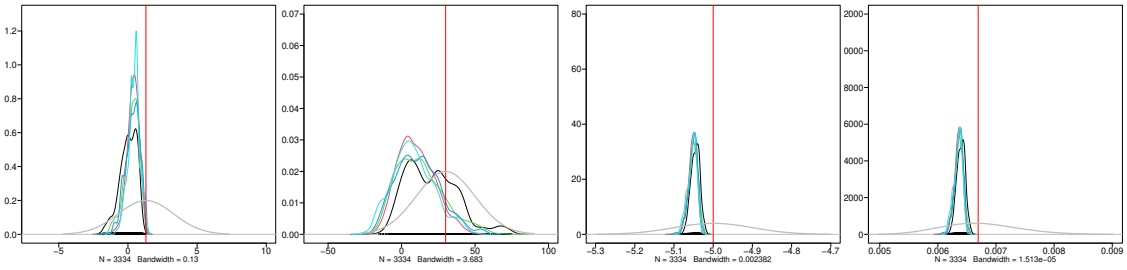
Finally, we forecast several steps ahead beyond the sample. We will use a similar methodology for the forecast, but it will be conditioned on all data. For these calculations, we assume that the Birth rate is constant and equal to the last observation; for the population size, we do a linear regression with the observations of the past five years, and then we use that to predict the population growth. For the effective vaccination rate, we consider three scenarios: optimistic (with an effective vaccination rate (EVR) of 95%), moderate (75% EVR), and pessimistic (50% EVR). The results of these forecasts are shown in Figures 3.13. Each row in this figure corresponds to each evaluated prior. In conclusion, all the estimates include the true value except for the first-time point in the pessimistic scenario with methods *MLETRAD* and *MAPTRAD*. These forecasts are also based on 1000 particles, as we explained above. Also, Figures 3.14 and 3.15 include residuals and studentized residuals. In these figures, we notice an increasing pattern for the residuals under the optimistic scenario and a decreasing pattern for the residuals under the pessimistic scenario. Furthermore, an increasing spread through time is common for all cases. These results, once again,

allow us to conclude about the robustness of the implemented Bayesian estimation method.

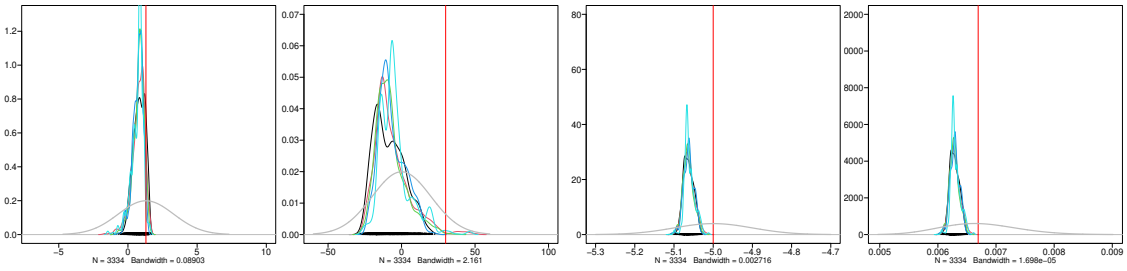
Figure 3.5: Posteriors estimations (colored lines) from pMCMC with initial values: β_0 =1.3, β_1 =30, and θ_r =-5 (first column β_0, second column β_1, third column θ_r, and fourth column ρ_r). Normal prior for the parameters β_0, β_1, θ_r. Each row corresponds to a priors (gray lines) as we specified in Table 3.2. The colored lines for the posteriors correspond to the selected proposals, see Table 3.2. The true parameter values is represented with a red line.



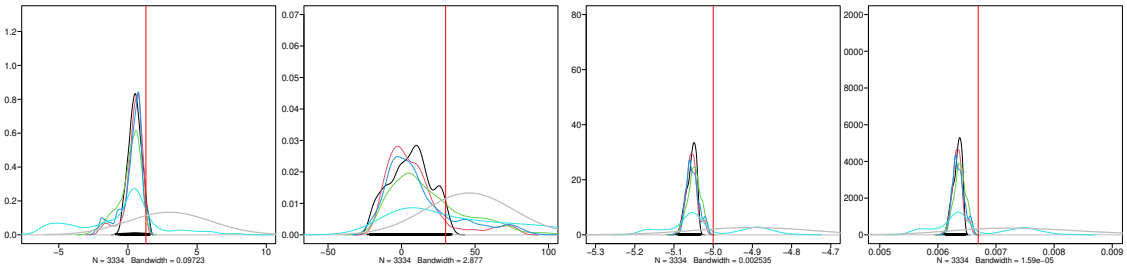
(a) β_0 (b) β_1 (c) θ_r (d) ρ_r



(e) (f) (g) (h)

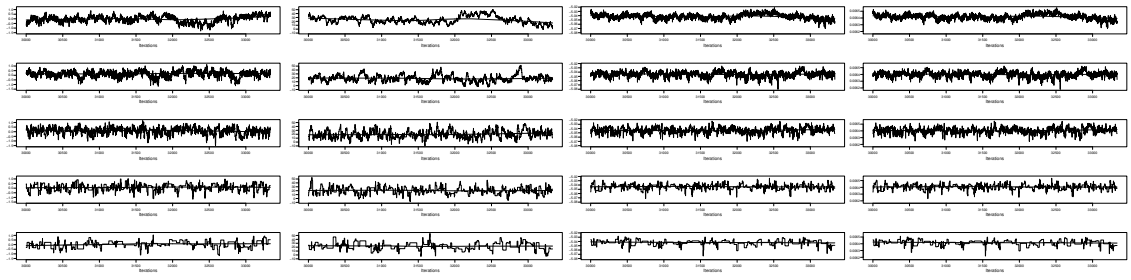


(i) (j) (k) (l)

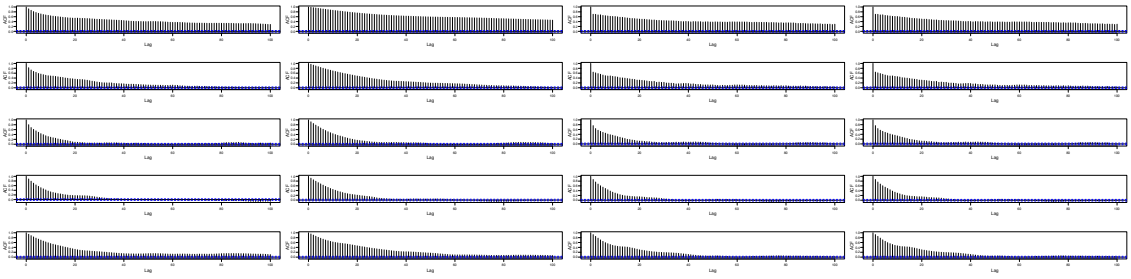


(m) (n) (o) (p)

Figure 3.6: Markov chains and estimated ACF with prior 1. Top: The state of the Markov chain. Each row correspond to the proposal specified in Table 3.2. Bottom: The estimated ACF of the Markov chain. Each row correspond to the proposal specified in Table 3.2.

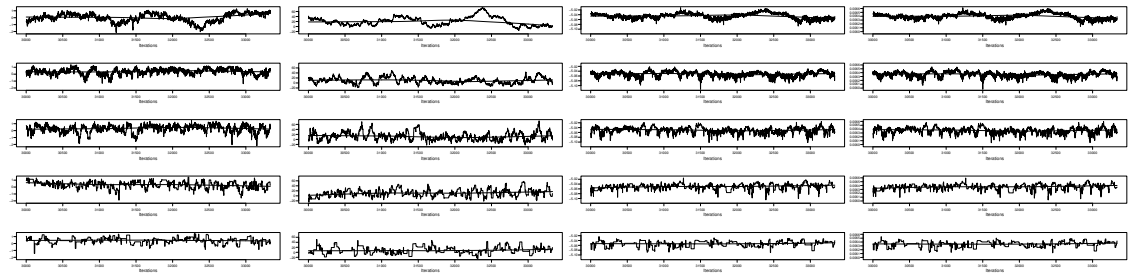


(a) β_0 (b) β_1 (c) θ_r (d) ρ_r

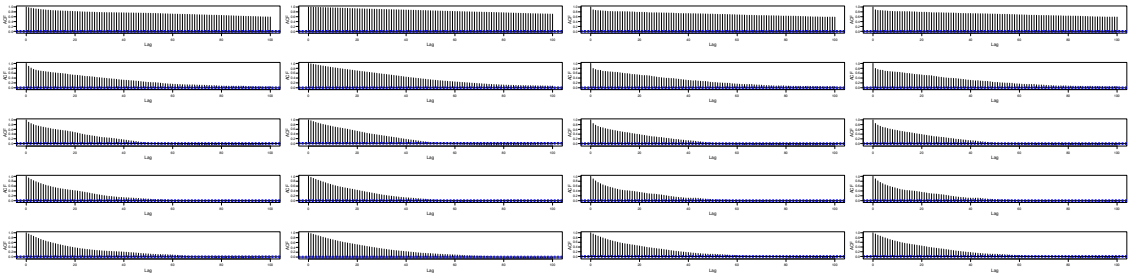


(e) (f) (g) (h)

Figure 3.7: Markov chains and estimated ACF with prior 2. Top: The state of the Markov chain. Each row correspond to the proposal specified in Table 3.2. Bottom: The estimated ACF of the Markov chain. Each row correspond to the proposal specified in Table 3.2.

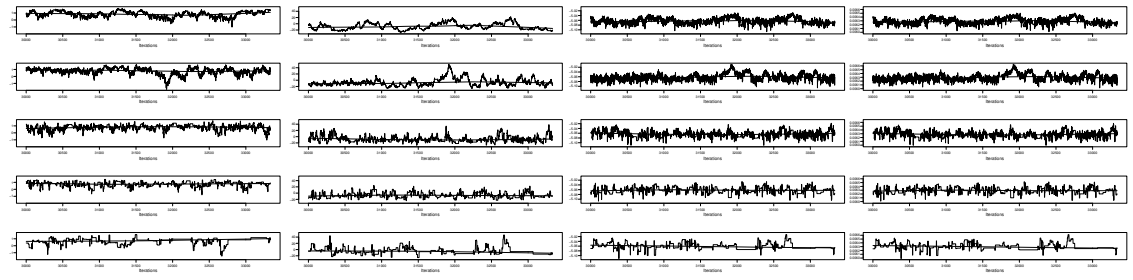


(a) β_0 (b) β_1 (c) θ_r (d) ρ_r

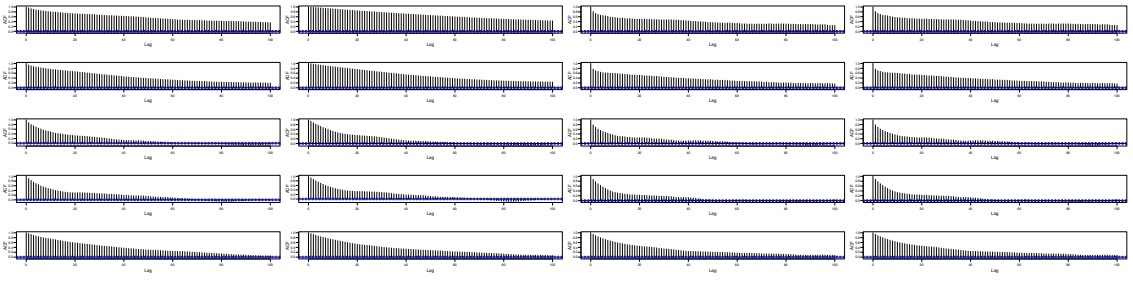


(e) (f) (g) (h)

Figure 3.8: Markov chains and estimated ACF with prior 3. Top: The state of the Markov chain. Each row correspond to the proposal specified in Table 3.2. Bottom: The estimated ACF of the Markov chain. Each row correspond to the proposal specified in Table 3.2.

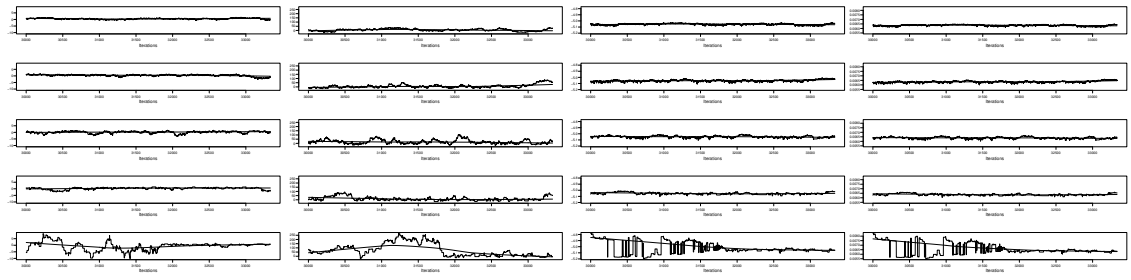


(a) β_0 (b) β_1 (c) θ_r (d) ρ_r

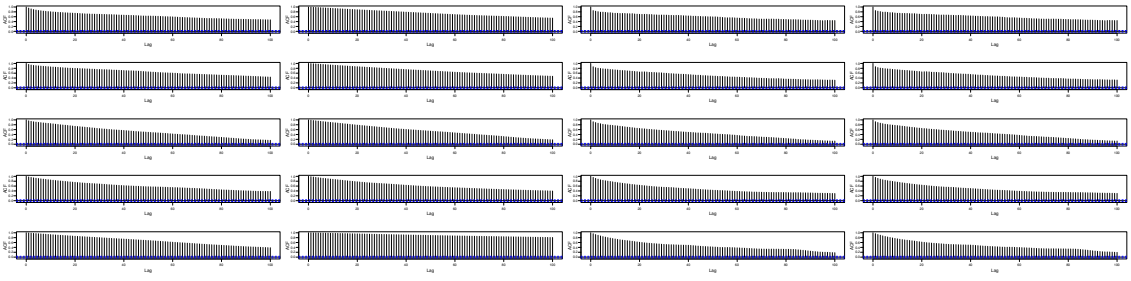


(e) (f) (g) (h)

Figure 3.9: Markov chains and estimated ACF with prior 4. Top: The state of the Markov chain. Each row correspond to the proposal specified in Table 3.2. Bottom: The estimated ACF of the Markov chain. Each row correspond to the proposal specified in Table 3.2.

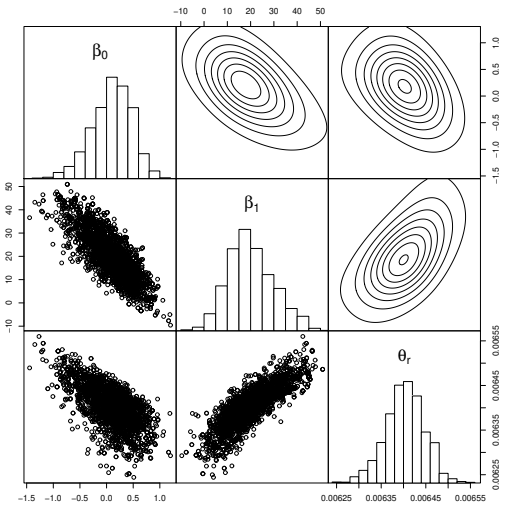


(a) β_0 (b) β_1 (c) θ_r (d) ρ_r

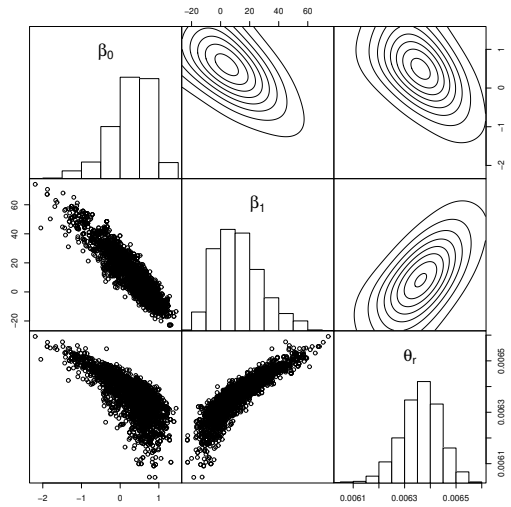


(e) (f) (g) (h)

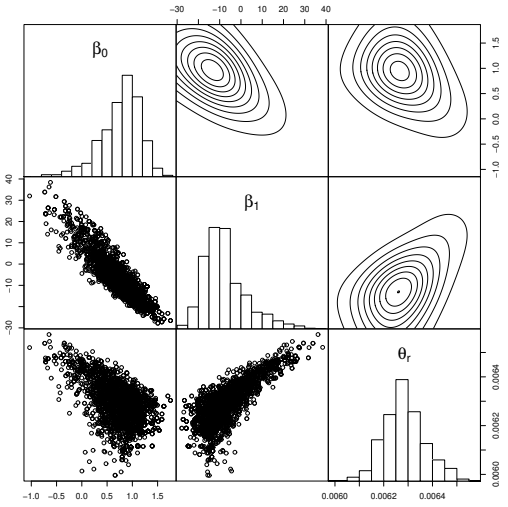
Figure 3.10: Parameters posterior estimation relationship with proposal (green) (Table 3.2). (a) Under prior 1, the Pearson correlation between $\hat{\beta}_0$ and $\hat{\beta}_1$ is 0.81, between $\hat{\beta}_0$ and $\hat{\theta}_r$ is 0.66, and between $\hat{\beta}_1$ and $\hat{\theta}_r$ is 0.85. (b) Under prior 2, the Pearson correlation between $\hat{\beta}_0$ and $\hat{\beta}_1$ is 0.92, between $\hat{\beta}_0$ and $\hat{\theta}_r$ is 0.76, and between $\hat{\beta}_1$ and $\hat{\theta}_r$ is 0.76. (c) Under prior 3, the Pearson correlation between $\hat{\beta}_0$ and $\hat{\beta}_1$ is 0.91, between $\hat{\beta}_0$ and $\hat{\theta}_r$ is 0.58, and between $\hat{\beta}_1$ and $\hat{\theta}_r$ is 0.78. (d) Under prior 4, the Pearson correlation between $\hat{\beta}_0$ and $\hat{\beta}_1$ is 0.96, between $\hat{\beta}_0$ and $\hat{\theta}_r$ is 0.86, and between $\hat{\beta}_1$ and $\hat{\theta}_r$ is 0.92.



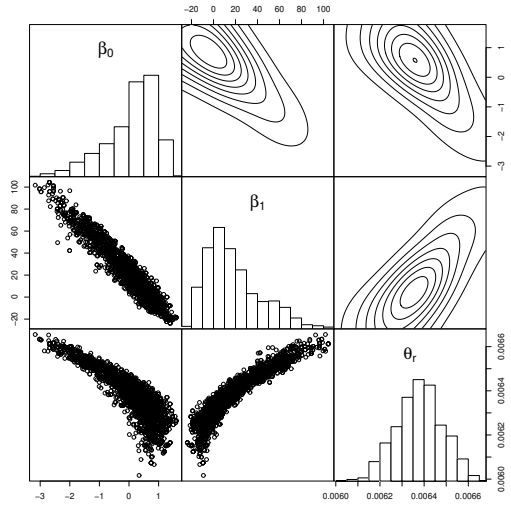
(a) With prior 1



(b) With prior 2

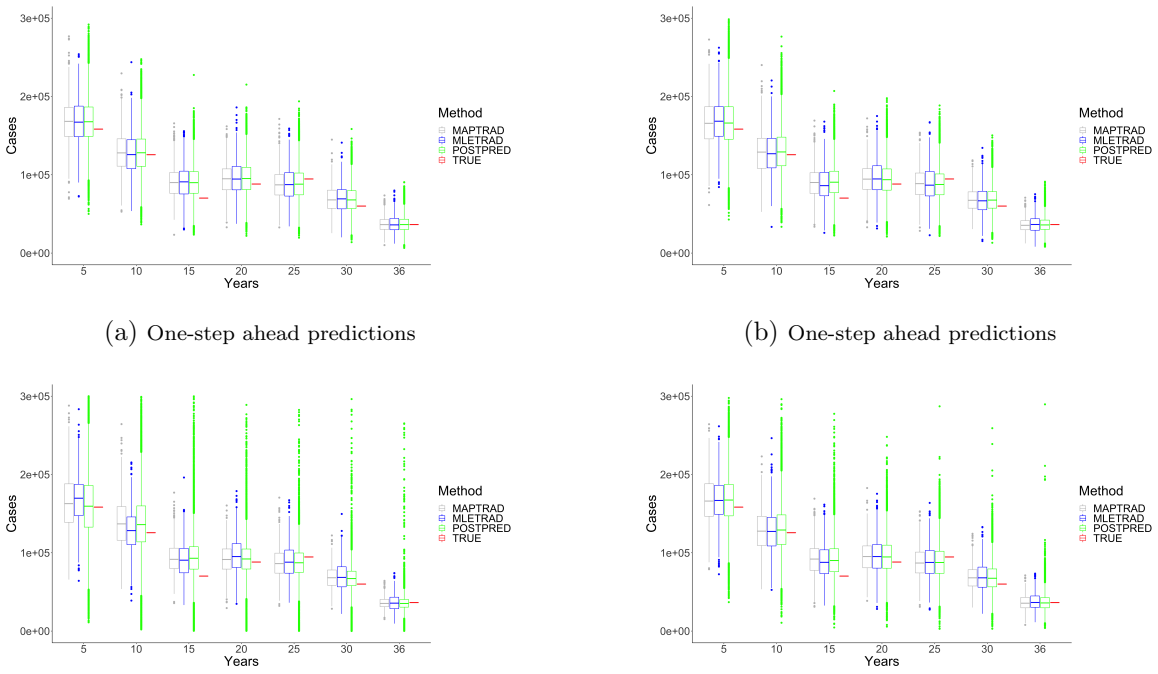


(c) With prior 3



(d) With prior 4

Figure 3.11: Comparison of the predictions inside sample (a) prior 1, (b) prior 2, (c) prior 3, and (d) prior 4. The posterior considered is the one under the case of proposal (green) (Table 3.2).



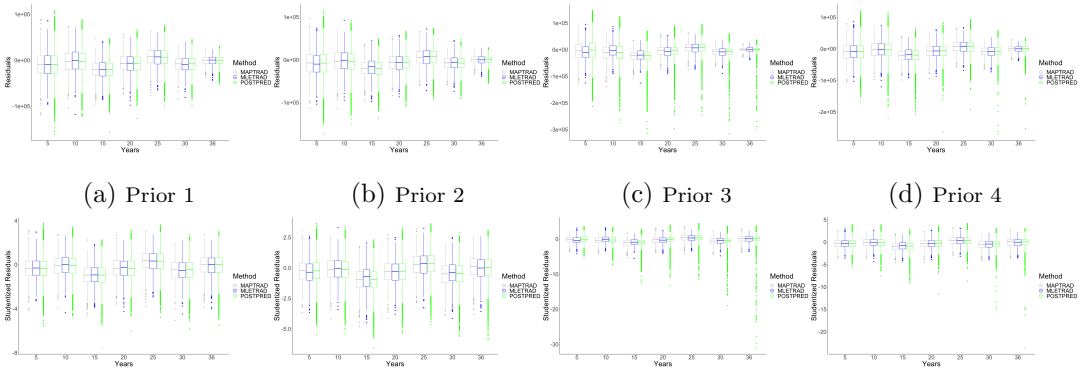
(a) One-step ahead predictions

(b) One-step ahead predictions

(c) One-step ahead predictions

(d) One-step ahead predictions

Figure 3.12: Comparison of the residuals inside sample (a) prior 1, (b) prior 2, (c) prior 3, and (d) prior 4 and studentized residuals inside sample (e) prior 1, (f) prior 2, (g) prior 3, and (h) prior 4. The posterior considered is the one under the case of proposal (green) (Table 3.2).



(a) Prior 1

(b) Prior 2

(c) Prior 3

(d) Prior 4

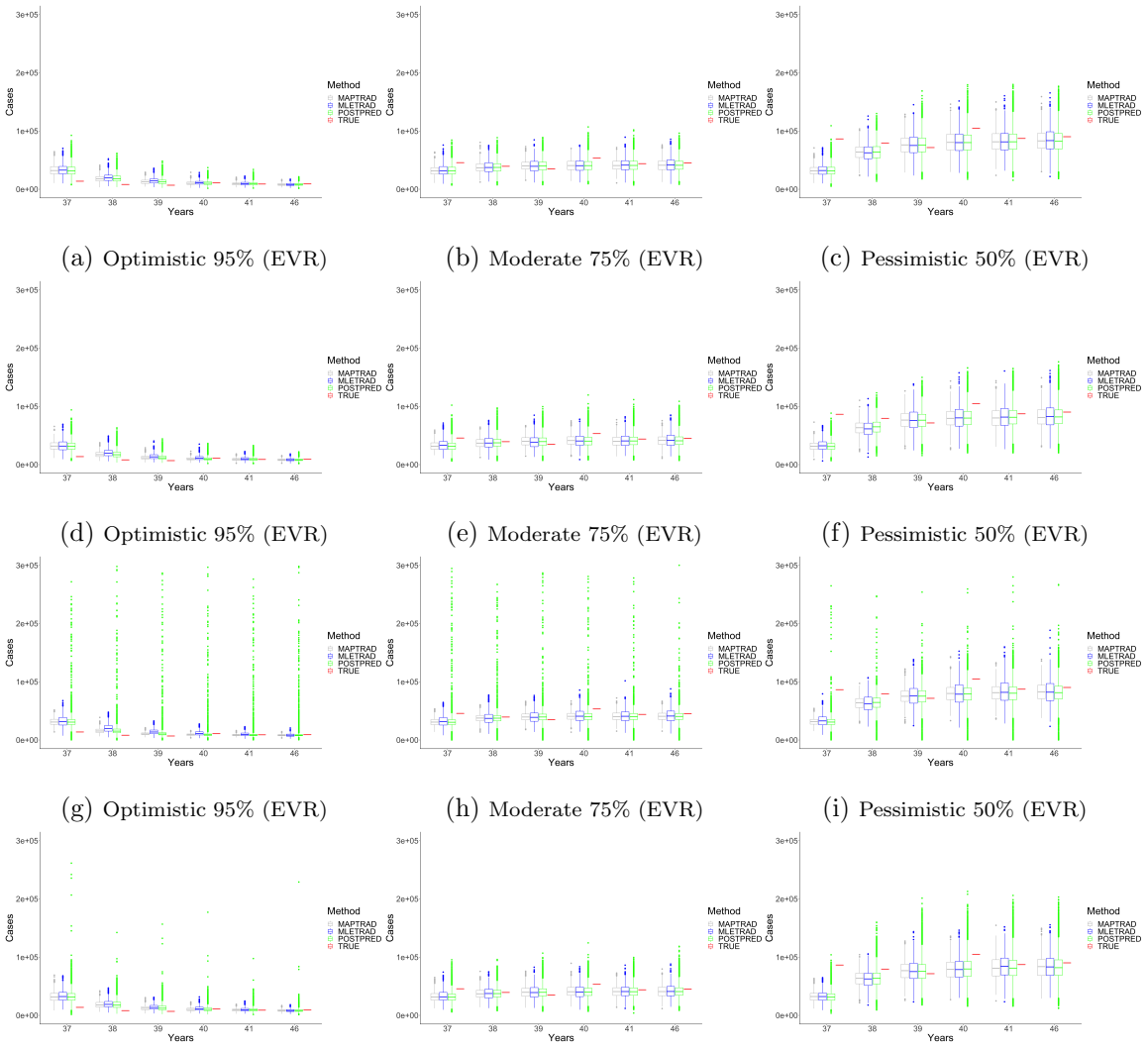
(e) Prior 1

(f) Prior 2

(g) Prior 3

(h) Prior 4

Figure 3.13: Comparison of the forecast using estimations from IF2 ($MLETRAD$), $MAPTRAD$ (taking the mean of the posterior distribution), and $POSTPRED$ (taking random draws from the posterior, respectively). Each row in the Figure correspond to each prior considered and posterior (green) (see Table 3.2). Three scenarios of the effective vaccination rate (EVR) are evaluated optimistic (95%), moderate (75%), and pessimistic (50%).



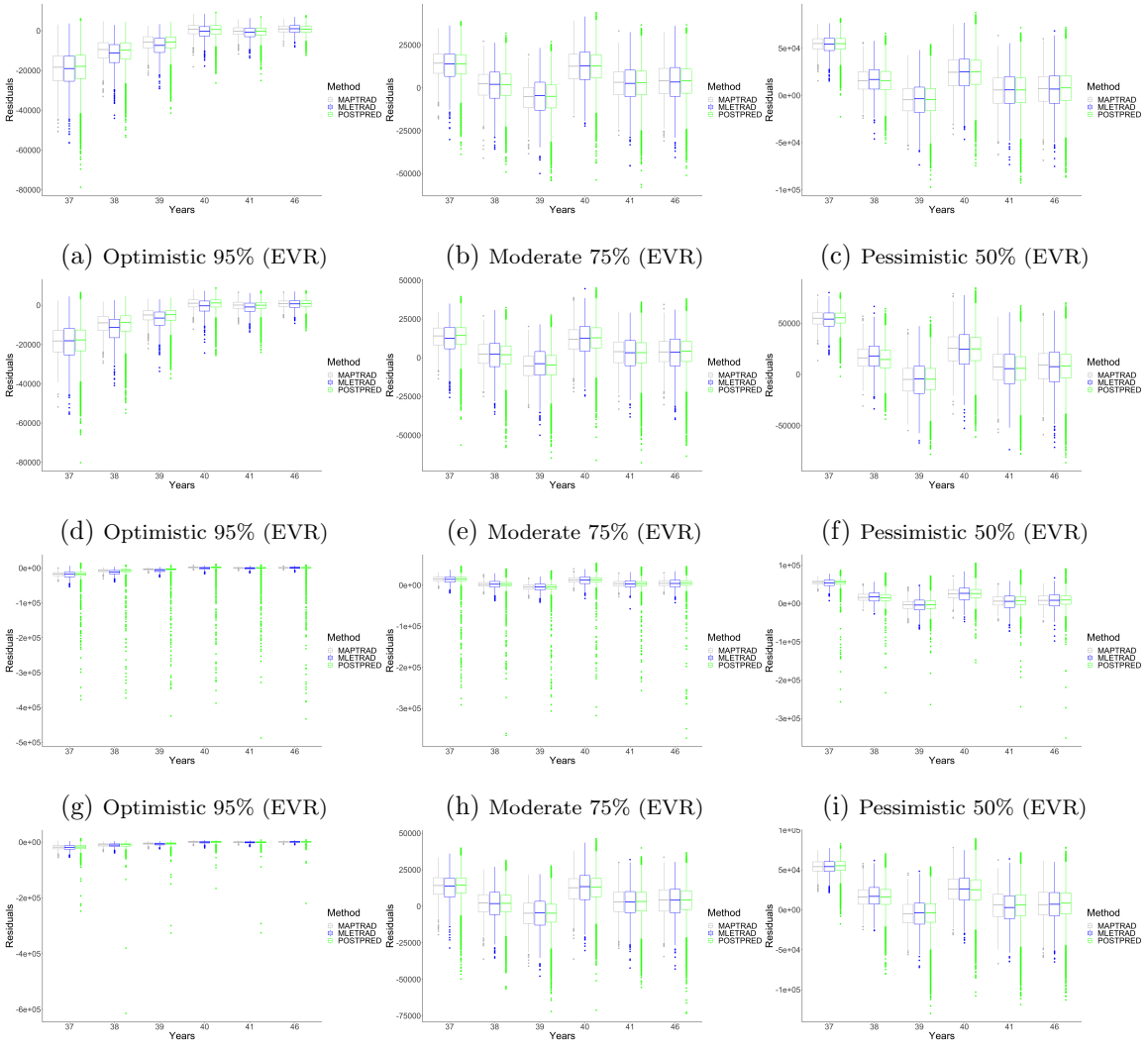
(a) Optimistic 95% (EVR) (b) Moderate 75% (EVR) (c) Pessimistic 50% (EVR)

(d) Optimistic 95% (EVR) (e) Moderate 75% (EVR) (f) Pessimistic 50% (EVR)

(g) Optimistic 95% (EVR) (h) Moderate 75% (EVR) (i) Pessimistic 50% (EVR)

(j) Optimistic 95% (EVR) (k) Moderate 75% (EVR) (l) Pessimistic 50% (EVR)

Figure 3.14: Forecast residuals using methods $MLETRAD$, $MAPTRAD$, and $POSTPRED$. Each row in the Figure correspond to each prior considered and posterior (green) (see Table 3.2). Three scenarios of the effective vaccination rate (EVR) are evaluated optimistic (95%), moderate (75%), and pessimistic (50%).



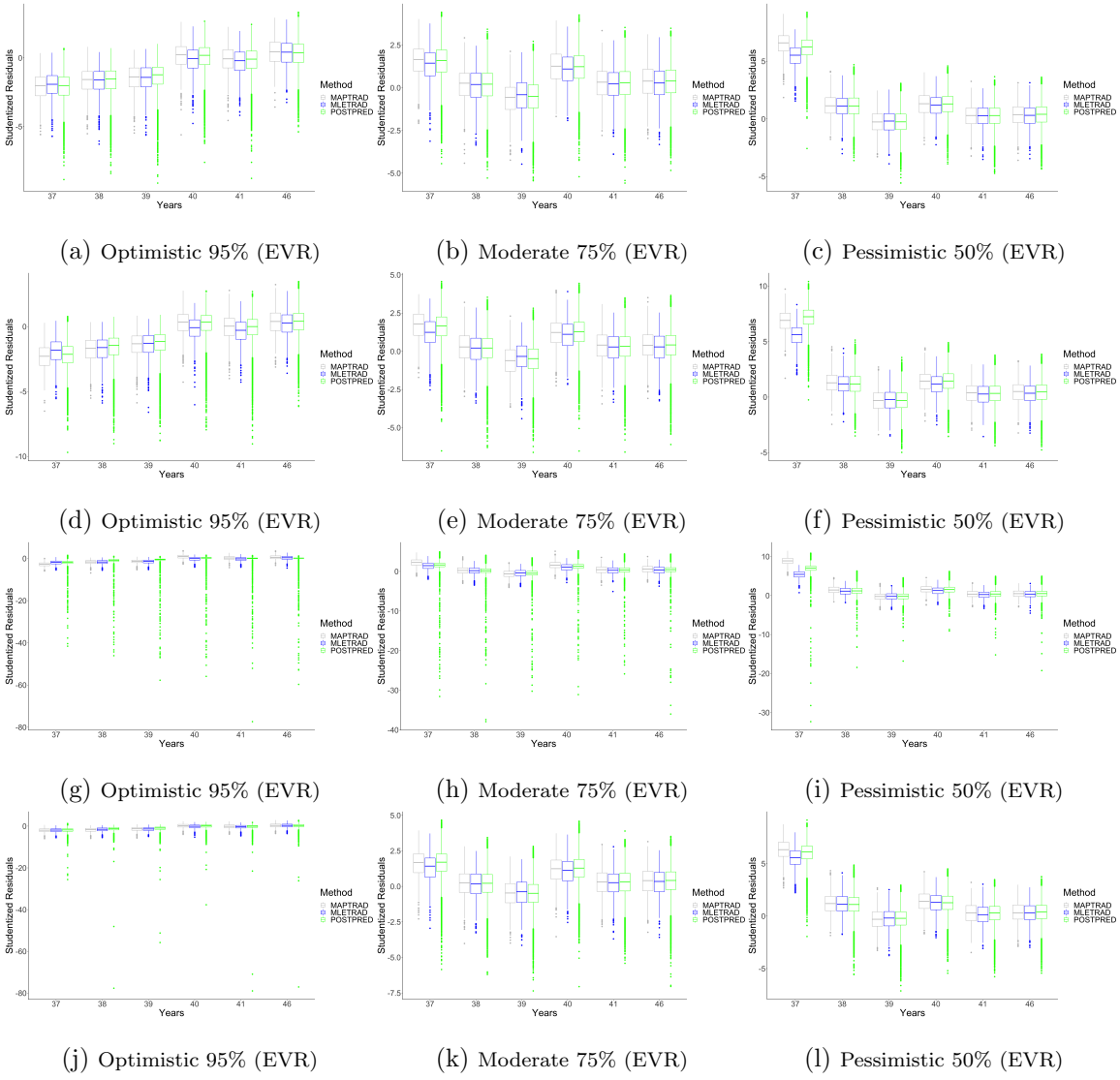
(a) Optimistic 95% (EVR) (b) Moderate 75% (EVR) (c) Pessimistic 50% (EVR)

(d) Optimistic 95% (EVR) (e) Moderate 75% (EVR) (f) Pessimistic 50% (EVR)

(g) Optimistic 95% (EVR) (h) Moderate 75% (EVR) (i) Pessimistic 50% (EVR)

(j) Optimistic 95% (EVR) (k) Moderate 75% (EVR) (l) Pessimistic 50% (EVR)

Figure 3.15: Forecast studentized residuals using methods $MLETRAD$, $MAPTRAD$, and $POSTPRED$. Each row in the Figure correspond to each prior considered and posterior (green) (see Table 3.2). Three scenarios of the effective vaccination rate (EVR) are evaluated optimistic (95%), moderate (75%), and pessimistic (50%).



(a) Optimistic 95% (EVR) (b) Moderate 75% (EVR) (c) Pessimistic 50% (EVR)

(d) Optimistic 95% (EVR) (e) Moderate 75% (EVR) (f) Pessimistic 50% (EVR)

(g) Optimistic 95% (EVR) (h) Moderate 75% (EVR) (i) Pessimistic 50% (EVR)

(j) Optimistic 95% (EVR) (k) Moderate 75% (EVR) (l) Pessimistic 50% (EVR)

3.3.2 Simulation Study Including the Parameter σ_e^2

Under this simulation study the parameter σ_e^2 is no longer fixed. Thus, our parameter vector to be estimated has four entries $\Theta = (\beta_0, \beta_1, \theta_r, \sigma_e^2)$. Again, we use particle MCMC (pMCMC) to facilitate Bayesian estimation and the IF2 as a support method. For the IF2, we used the same conditions described in section 3.3.1. In Figure 3.16, we plot the estimations from IF2.

For the Bayesian inference, this study considers similar conditions for the parameter β_0, β_1, and θ_r as those considered in the previous section, i.e., normal priors and normal random walk proposals. For the parameter σ_e^2, we consider inverse gamma priors and we define the transformation parameter $\tau_e^2 = \log(\sigma_e^2)$ to avoid restrictions on the exploration of the parameters space. The prior calculation of the parameter τ_e^2 is given as follows.

Prior transformation for τ_e^2. $\sigma_e^2 \sim IG(a, b)$.

$$\begin{aligned} \tau_e^2 &= \log(\sigma_e^2), \\ \sigma_e^2 &= exp\{\tau_e^2\}. \\ f_{\tau_e^2}(\tau_e^2) &= f_{\sigma_e^2}\left(exp\{\tau_e^2\}\right)\left|\frac{d}{d\tau_e^2}exp\{\tau_e^2\}\right|, \\ f_{\tau_e^2}(\tau_e^2) &= g_{\sigma_e^2}\left(exp\{-\tau_e^2\}\right)exp\{-\tau_e^2\}, \end{aligned} \tag{3.9}$$

where $g_{\sigma_e^2}(\cdot) \sim G(a, b)$. Thus, the proposal for τ_e^2 is also a normal random walk. The specifications of the priors and the proposals considered in this study are described in Table 3.3. According to these settings the results from pMCMC are shown in figures 3.17 - 3.21. By combining the three selected priors and the five selected proposals, we show the kernel density estimates in Figure 3.17. Each row represents the results of the estimation of the posterior densities (colored lines, the colored lines correspond to the selected proposals, see Table 3.3) associated with each fixed prior (gray line).

The red line represents the true parameter value. These results again do not show substantial changes in the posterior densities with respect to the choice of the priors for the parameters β_0, β_1, and θ_r, which allows us to conclude the robustness of the Bayesian estimation. However, we can not conclude the same for the posterior estimate of the parameter τ_e^2 (σ_e^2). In this case, the posterior is highly sensitive to the prior choice. In Figures 3.21 we show the relationship between the estimated posterior densities. In these plots, we observe a high correlation between β_0 and β_1. Also, a low correlation is observed between τ_e^2 and the other parameters.

Figures 3.22 and 3.24 show the results of the one-step ahead predictions and forecast under the three scenerios (pessimistic, moderate and optimistic) described in the previous section. Figures 3.22 exhibit an increasing spread of the predictions as the mean of the prior for σ_e^2 increases. We can see the same behavior more clearly, for the forecast. These results mimic the ones obtained by Eilertson *et al.* (2019), the likelihood is insensitive to the σ_e^2 parameter, and therefore the posterior is sensitive to its priors. Given these results, the parameter σ_e^2 is fixed in the next section, where we analyzed the model in the context of real data. Also, in Figures 3.23, 3.25, and 3.26, we include the residuals and the studentized residuals.

Table 3.3: Specifications of the Prior and Proposal Distributions

			β_0	β_1	θ_r		σ_e^2
Prior 1	Normal	μ	1.3	30	-5	Beta	$(5, 1)$
	(grey line)	σ^2	2^2	30^2	0.1^2		
Prior 2	Normal	μ	1.3	30	-5	Beta	$(5, 4)$
	(grey line)	σ^2	2^2	30^2	0.1^2		
Prior 3	Normal	μ	3	46	-4.9	Beta	$(5, 9)$
	(grey line)	σ^2	3^2	30^2	1.5^2		

		Standard deviation (ϵ_j) proposal distribution			
		β_0	β_1	θ_r	τ_e^2
Posterior					
—	$j = 1$	0.05	0.5	0.005	0.005
—	$j = 2$	0.1	1	0.01	0.01
—	$j = 3$	0.5	5	0.05	0.05
—	$j = 4$	1	10	0.1	0.1
—	$j = 4$	1.5	15	0.15	0.15

Figure 3.16: Results from IF2 with initial values: β_0 =1.3, β_1 =30, θ_r =-5, and σ_e^2 =0.5

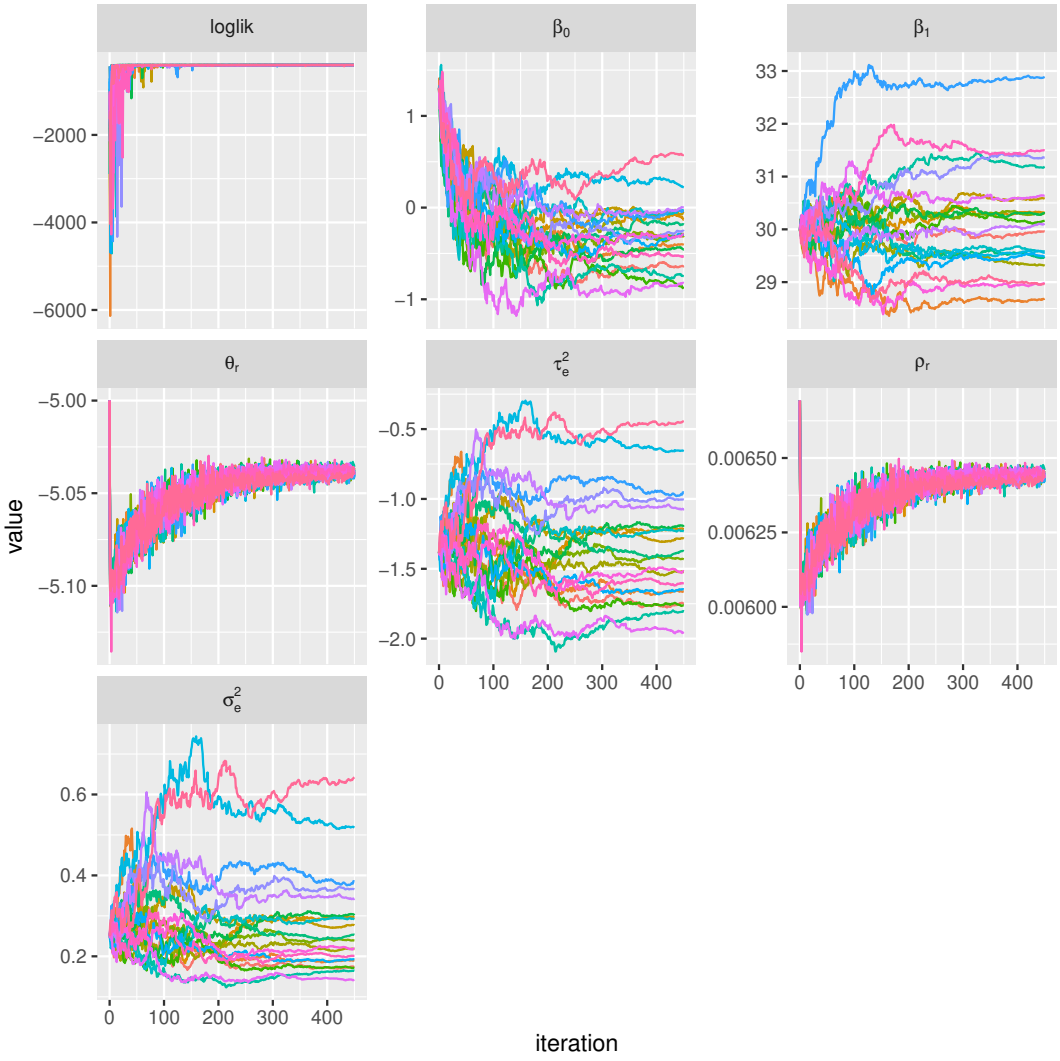
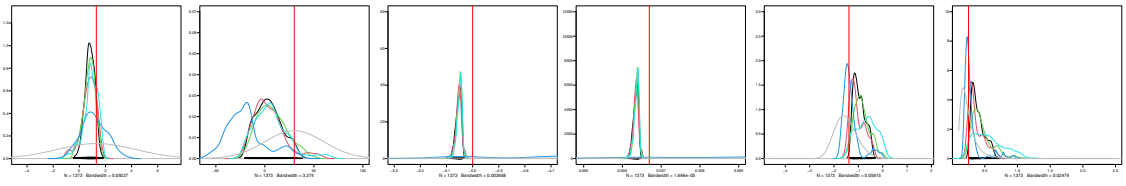
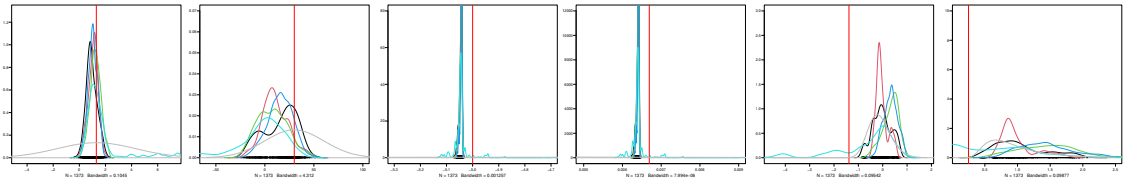


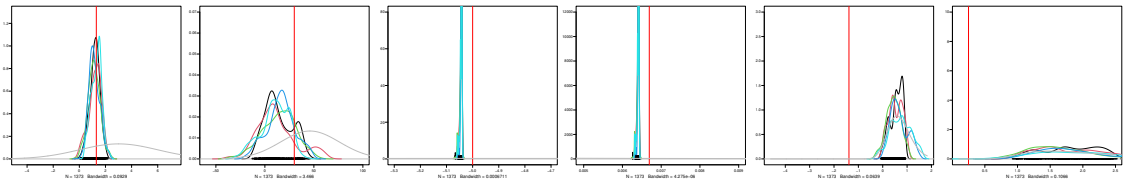
Figure 3.17: Posteriors estimations (colored lines) from pMCMC with initial values: β_0 =1.3, β_1 =30, θ_r =-5 and σ_e^2 =0.5 (first column β_0, second column β_1, third column θ_r, fourth column ρ_r, fifth column, sixth column τ_e^2, and seventh column σ_e^2). Normal priors for the parameters β_0, β_1, θ_r, and inverse gamma prior for σ_e^2. Each row corresponds to a priors (gray lines) as we specified in Table 3.3. The colored lines for the posteriors correspond to the selected proposals, see Table 3.3. The true parameter values is represented with a red line.



(a) β_0 (b) β_1 (c) θ_r (d) ρ_r (e) τ_e^2 (f) σ_e^2

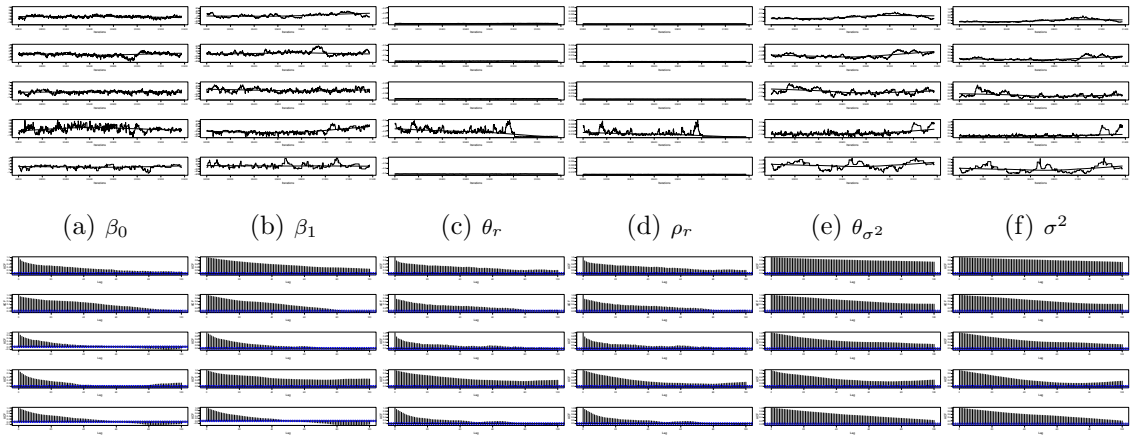


(g) (h) (i) (j) (k) (l)



(m) (n) (o) (p) (q) (r)

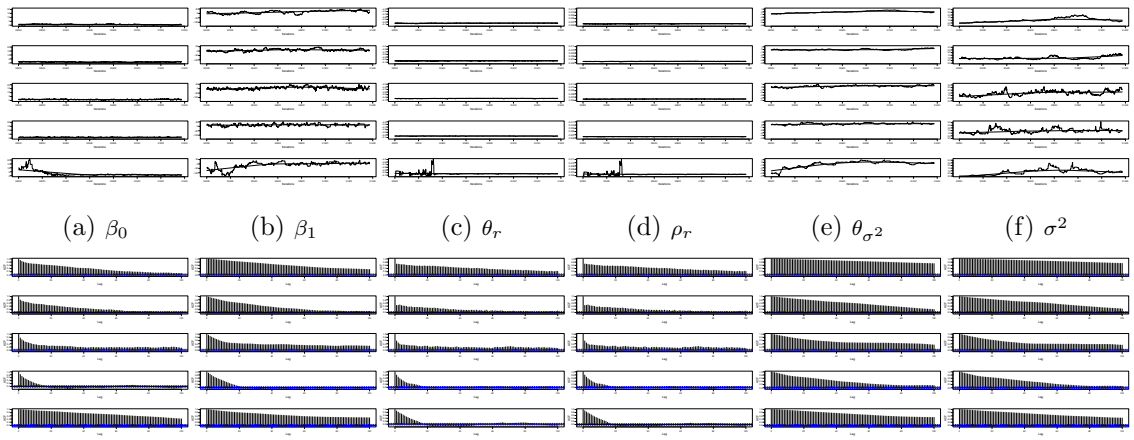
Figure 3.18: Markov chains and estimated ACF with prior 1. Top: The state of the Markov chain. Each row correspond to the proposal specified in Table 3.3. Bottom: The estimated ACF of the Markov chain. Each row correspond to the proposal specified in Table 3.3.



(a) β_0 (b) β_1 (c) θ_r (d) ρ_r (e) θ_{σ^2} (f) σ^2

(g) (h) (i) (j) (k) (l)

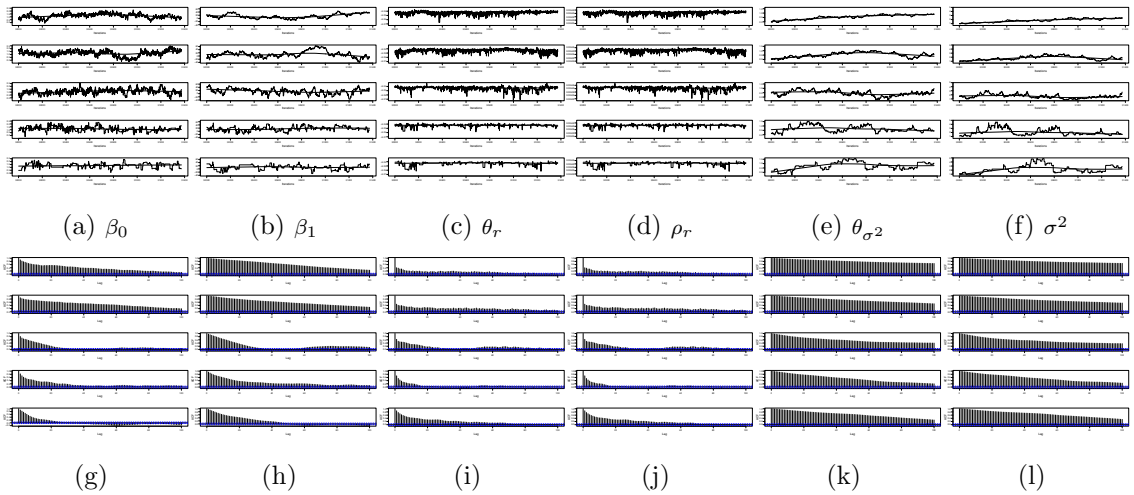
Figure 3.19: Markov chains and estimated ACF with prior 2. Top: The state of the Markov chain. Each row correspond to the proposal specified in Table 3.3. Bottom: The estimated ACF of the Markov chain. Each row correspond to the proposal specified in Table 3.3.



(a) β_0 (b) β_1 (c) θ_r (d) ρ_r (e) θ_{σ^2} (f) σ^2

(g) (h) (i) (j) (k) (l)

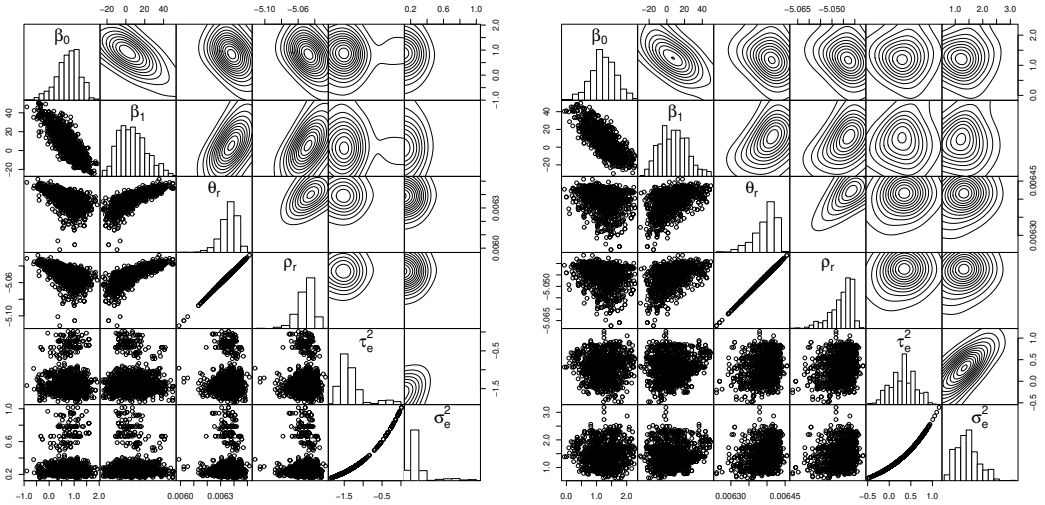
Figure 3.20: Markov chains and estimated ACF with prior 3. Top: The state of the Markov chain. Each row correspond to the proposal specified in Table 3.3. Bottom: The estimated ACF of the Markov chain. Each row correspond to the proposal specified in Table 3.3.



(a) β_0 (b) β_1 (c) θ_r (d) ρ_r (e) θ_{σ^2} (f) σ^2

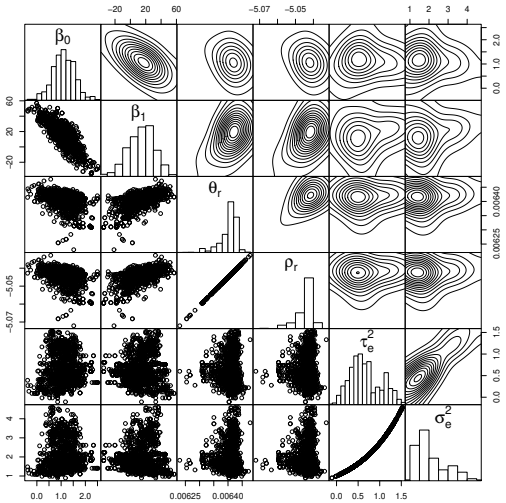
(g) (h) (i) (j) (k) (l)

Figure 3.21: Parameters posterior estimation relationship with proposal (green) (Table 3.3). (a) Under prior 1, the Pearson correlation between $\hat{\beta}_0$ and $\hat{\beta}_1$ is 0.83, between $\hat{\beta}_0$ and $\hat{\theta}_r$ is 0.51, and between $\hat{\beta}_1$ and $\hat{\theta}_r$ is 0.79. (b) Under prior 2, the Pearson correlation between $\hat{\beta}_0$ and $\hat{\beta}_1$ is 0.79, between $\hat{\beta}_0$ and $\hat{\theta}_r$ is 0.26, and between $\hat{\beta}_1$ and $\hat{\theta}_r$ is 0.54. (c) Under prior 3, the Pearson correlation between $\hat{\beta}_0$ and $\hat{\beta}_1$ is 0.83, between $\hat{\beta}_0$ and $\hat{\theta}_r$ is 0.46, and between $\hat{\beta}_1$ and $\hat{\theta}_r$ is 0.64.



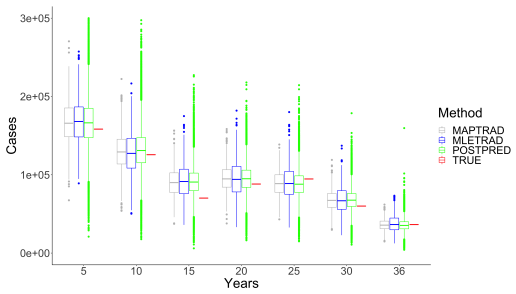
(a) With prior 1

(b) With prior 2

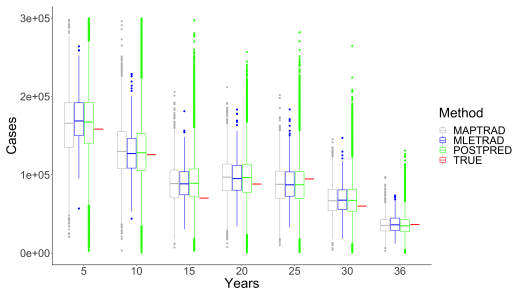


(c) With prior 3

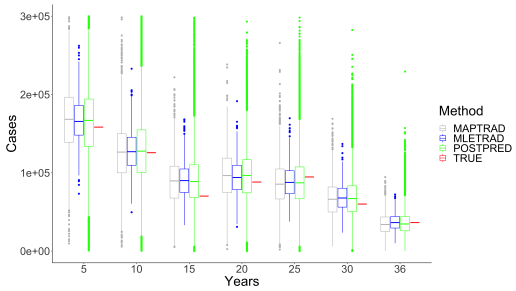
Figure 3.22: Comparison of the predictions inside sample (a) prior 1, (b) prior 2, and (c) prior 3. The posterior considered is the one under the case of proposal (green) (Table 3.3).



(a) One-step ahead predictions

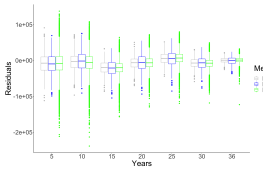


(b) One-step ahead predictions

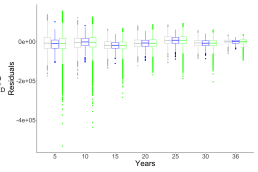


(c) One-step ahead predictions

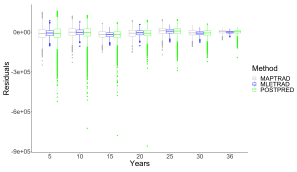
Figure 3.23: Comparison of the residuals inside sample (a) prior 1, (b) prior 2, (c) prior 3, and studentized residuals inside sample (e) prior 1, (f) prior 2, (g) prior 3. The posterior considered is the one under the case of proposal (green) (Table 3.3).



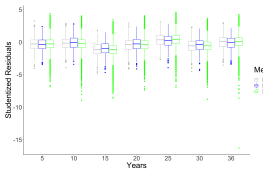
(a) Prior 1



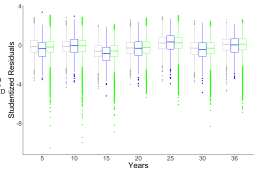
(b) Prior 2



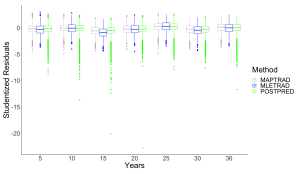
(c) Prior 3



(d) Prior 1

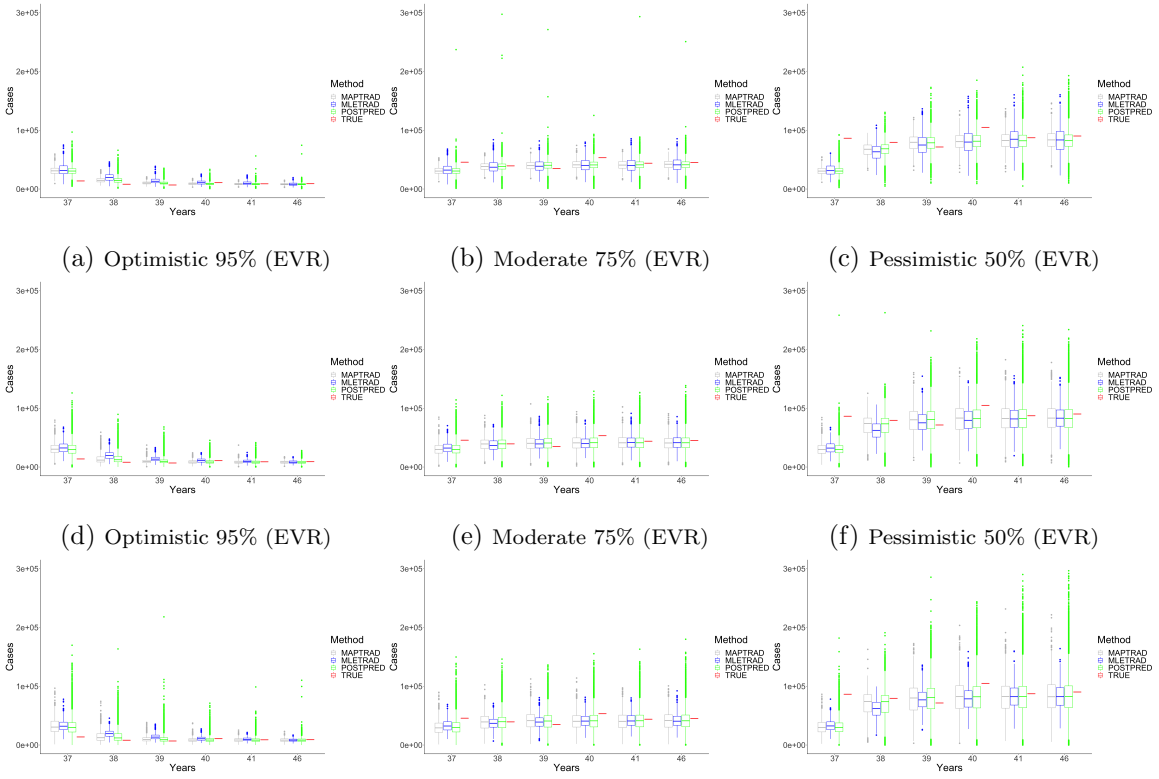


(e) Prior 2



(f) Prior 3

Figure 3.24: Comparison of the forecast using estimations from IF2 $MLETRAD$, $MAPTRAD$ (taking the mode of the posterior distribution), and $POSTPRED$ (taking random draws from the posterior, respectively). Each row in the Figure correspond to each prior considered and posterior (green) (see Table 3.3). Three scenarios of the effective vaccination rate (EVR) are evaluated optimistic (95%), moderate (75%), and pessimistic (50%).

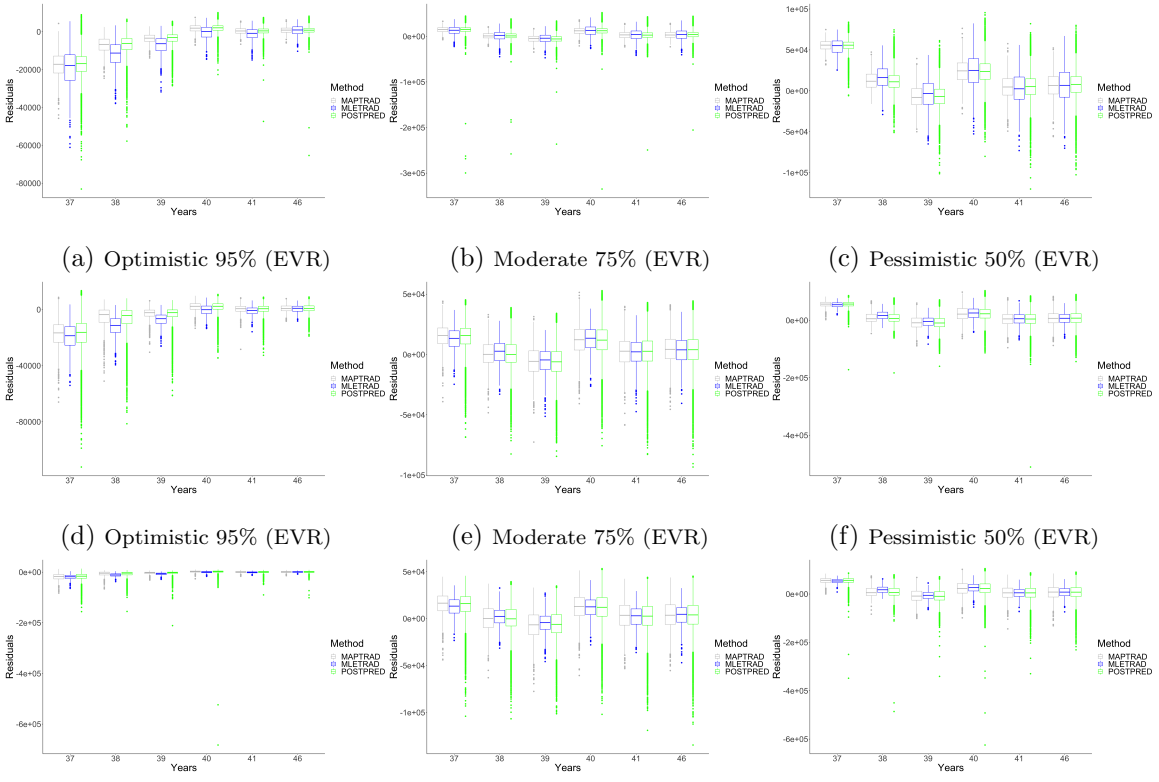


(a) Optimistic 95% (EVR) (b) Moderate 75% (EVR) (c) Pessimistic 50% (EVR)

(d) Optimistic 95% (EVR) (e) Moderate 75% (EVR) (f) Pessimistic 50% (EVR)

(g) Optimistic 95% (EVR) (h) Moderate 75% (EVR) (i) Pessimistic 50% (EVR)

Figure 3.25: Forecast residuals using methods $MLETRAD$, $MAPTRAD$, and $POSTPRED$. Each row in the Figure correspond to each prior considered and posterior (green) (see Table 3.2). Three scenarios of the effective vaccination rate (EVR) are evaluated optimistic (95%), moderate (75%), and pessimistic (50%).

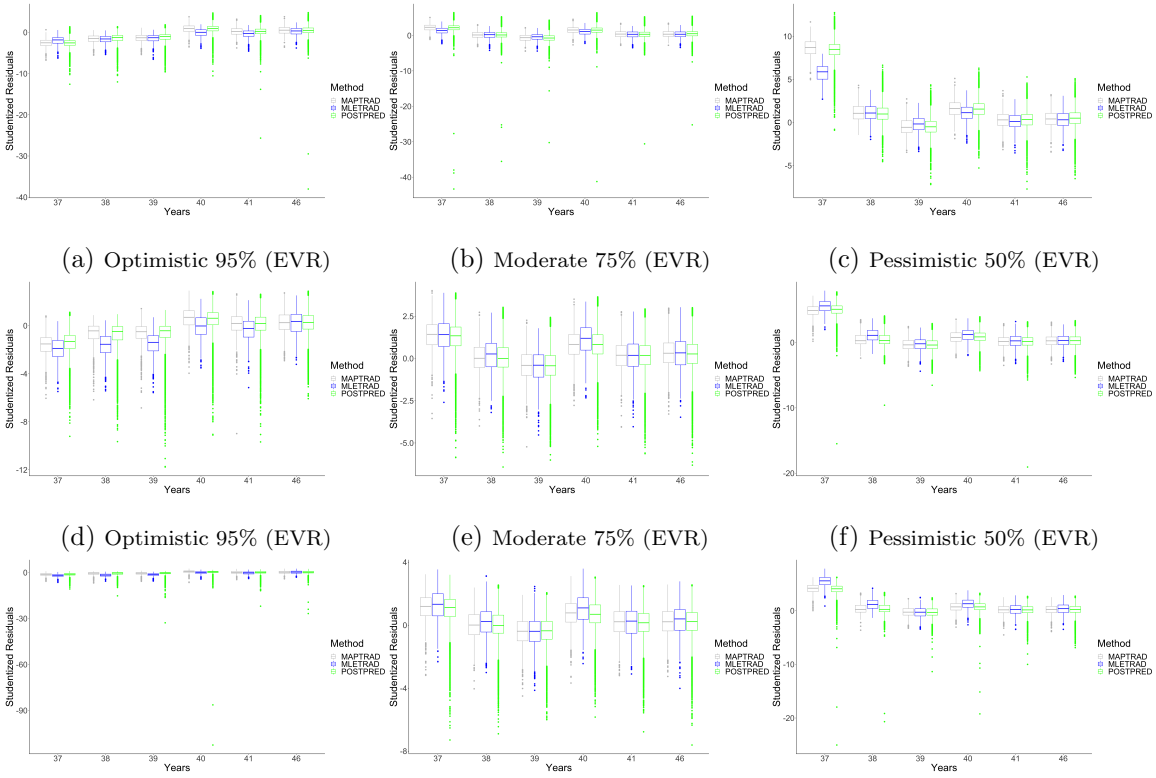


(a) Optimistic 95% (EVR) (b) Moderate 75% (EVR) (c) Pessimistic 50% (EVR)

(d) Optimistic 95% (EVR) (e) Moderate 75% (EVR) (f) Pessimistic 50% (EVR)

(g) Optimistic 95% (EVR) (h) Moderate 75% (EVR) (i) Pessimistic 50% (EVR)

Figure 3.26: Forecast studentized residuals using methods $MLETRAD$, $MAPTRAD$, and $POSTPRED$. Each row in the Figure correspond to each prior considered and posterior (green) (see Table 3.2). Three scenarios of the effective vaccination rate (EVR) are evaluated optimistic (95%), moderate (75%), and pessimistic (50%).



(a) Optimistic 95% (EVR) (b) Moderate 75% (EVR) (c) Pessimistic 50% (EVR)

(d) Optimistic 95% (EVR) (e) Moderate 75% (EVR) (f) Pessimistic 50% (EVR)

(g) Optimistic 95% (EVR) (h) Moderate 75% (EVR) (i) Pessimistic 50% (EVR)

3.4 Case Studies

In this section, we fit the model (3.1-3.5) to observed data. We focus on the annual measles reports in India (IND), China (CHN), and United Arab Emirates (ARE) between 1980 and 2014. The data also include the yearly population size, the number of births, and the effective vaccination rate as covariates; Figures 3.28a, 3.33a, and 3.38a report these data. We perform a Bayesian analysis of this data with the form of the priors as previously, and we will use IF2 to help us select appropriate priors, so this is a kind of empirical Bayes. These two algorithms pMCMC and IF2 are described in sections 2.2.2 and 2.2.3, respectively.

We fit the model in (3.1-3.5) using the results of the estimation of the parameters obtained in Eilertson *et al.* (2019) as starting values for the parameters to do a preliminary analysis. Also, the simulation results in the previous sections help us select priors and proposals. Furthermore, we use data from 12 countries (we only report the results for the countries mentioned before). Thus we propose the following steps:

1. We take some initial values, run IF2 for those, and do that for a couple of different initial values, and see if they agree.

2. Given the data for any country, we do a search of the parameters β_0, β_1, and θ_r to find appropriate initial values for S_0 in (3.1) such that the prediction of the number of cases $\hat{C}_1$ be reasonably close to the observation C_1. To do that, we set a grid of K uniform random values around the estimates from step 1, $[\hat{\theta}_0^{(1)} - \delta, \hat{\theta}_0^{(1)} + \delta]$ (the super script is used to show that the estimation comes from step 1), for each of three parameters. Then, we calculate $\hat{C}_{1k}$, $k = 1, 2, 3, ..., K$, and keep the one whose $C_1 - \hat{C}_{1k}$ is the minimum.

3. Given the parameter values in step 2, we rerun IF2. From this result, we set starting values for the parameter vector Θ and we set priors, see Table 3.4.

4. We run pMCMC by using the results in the previous step. The proposal is selected by following the results from simulations in the earlier sections; these are summarized in Table 3.4.

We split the results of this application by country. Figures 3.28b, 3.33b, and 3.38b show the estimates from IF2. The conditions for IF2 are 20000 particles, 450 iterations, and a cooling parameter of 0.7. The three cases show convergence. For the Bayesian estimation, we simulate independent chains with 50^4 particles and 10^5 iterations. For the three cases, we burned the first 30^4 iterations and thin parameters 20, 30, and 10 for IND, CHN, and ARE, respectively. Figures in 3.28 include density plots (prior, gray line, and posterior colored lines), ACF plots, and trace plots for India. Similarly, Figures in 3.33 and 3.38 include the same results for China and United Arab Emirates, respectively.

In Figures 3.29, 3.34, and 3.39, we can see the relationship of the estimated parameters for IND, CHN, and ARE, respectively. For the ARE country particularly, we can identify a non-linear dependency between the estimations.

To evaluate the quality of these estimators, we chose one of the posterior estimates to calculate one-step ahead predictions and forecasts. From the estimated ACF of the Markov chain, we choose the posterior related to the proposal whose ACF is close to uncorrelated. Thus, we choose the posterior (red), (black), and (blue), for IND, CHN, and ARE, respectively (see Table 3.4). In Figures 3.30, we summarized one-step ahead predictions and forecast for the three scenarios, optimistic (with an effective vaccination rate (EVR) of 95%), moderate (75% EVR), and pessimistic (50% EVR). For the covariates, we assume that the Birth rate is constant and equal to the last

observation; for the population size, we do a linear regression with the observations of the past five years, and then we use that to predict the population growth. Also, we compare the prediction capacity with the estimations from IF2 and pMCMC. From pMCMC outcomes, we perform MAP predictions, *MAPTRAD*, and *POSTPRED*, as we described in section 3.3.1. We display the results for some time points about the one-step ahead predictions for better visualization. In general, the three methods provide predictions that cover the number of infected cases. In other words, the IF2 (*MLETRAD*) forecasts a fewer number of infecteds than the other two ways. This behavior is common under the three scenarios.

Figures in 3.35 and 3.40 include prediction outcomes for the countries CHN and ARE, respectively. In the case of CHN, we obtain more similar predictions under three methods for both one-step ahead predictions and forecasts. Regarding ARE, the (*MLETRAD*) forecasts less infecteds than the other two methods. However, there is an increasing spread of the forecast under the three scenarios as the forecast horizon increase.

Finally, Figures 3.31, 3.36, and 3.41 include both the residuals and standarized residuals.

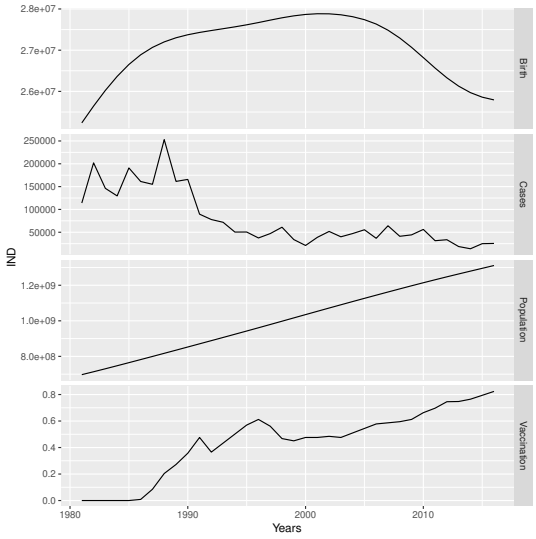
Table 3.4: Specifications of the Prior and Proposal Distributions

Priors			β_0	β_1	θ_r
IND	Normal	μ	-4.89	35	-5.03
	(grey line)	σ^2	3^2	20^2	0.5^2
CHN	Normal	μ	-3	32	-3.3
	(grey line)	σ^2	0.9^2	3^2	0.1^2
ARE	Normal	μ	-7.06	75.18	-1.8
	(grey line)	σ^2	0.1^2	1^2	0.1^2

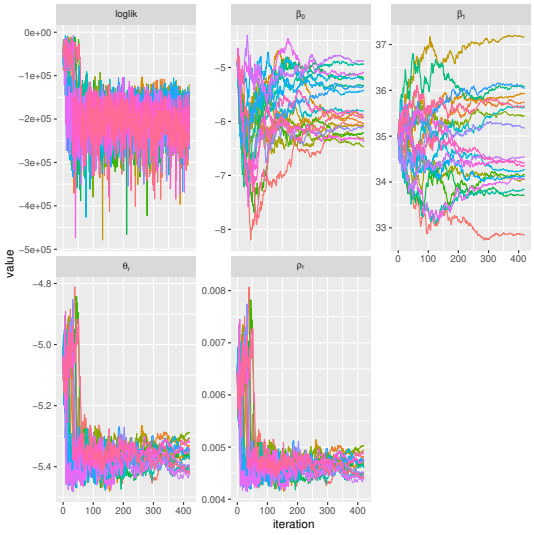
		Standard deviation (ϵ_j) proposal distribution		
Posterior		β_0	β_1	θ_r
IND				
—	$j = 1$	0.02	0.02	0.01
—	$j = 2$	0.75	5	0.015
—	$j = 3$	1.5	10	0.03
—	$j = 4$	2	15	0.06
—	$j = 5$	3	20	0.12
CHN				
—	$j = 1$	0.15	0.5	0.012
—	$j = 2$	0.3	1	0.025
—	$j = 3$	0.6	2	0.05
—	$j = 4$	0.9	3	0.075
—	$j = 5$	1.5	5	0.125
ARE				
—	$j = 1$	0.02	0.02	0.01
—	$j = 2$	0.04	0.08	0.02
—	$j = 3$	0.16	0.5	0.04
—	$j = 4$	0.3	2	0.08
—	$j = 5$	0.5	5	0.1

Country of India (IND)

Figure 3.27: (a) Data cases from country of India refer to the number of births, number of reported measles cases, population size, and the effective vaccination rate. (b) Parameter estimation by using IF2 20 independent chains are evaluated, 420 iterations and a cooling parameter of 0.7.

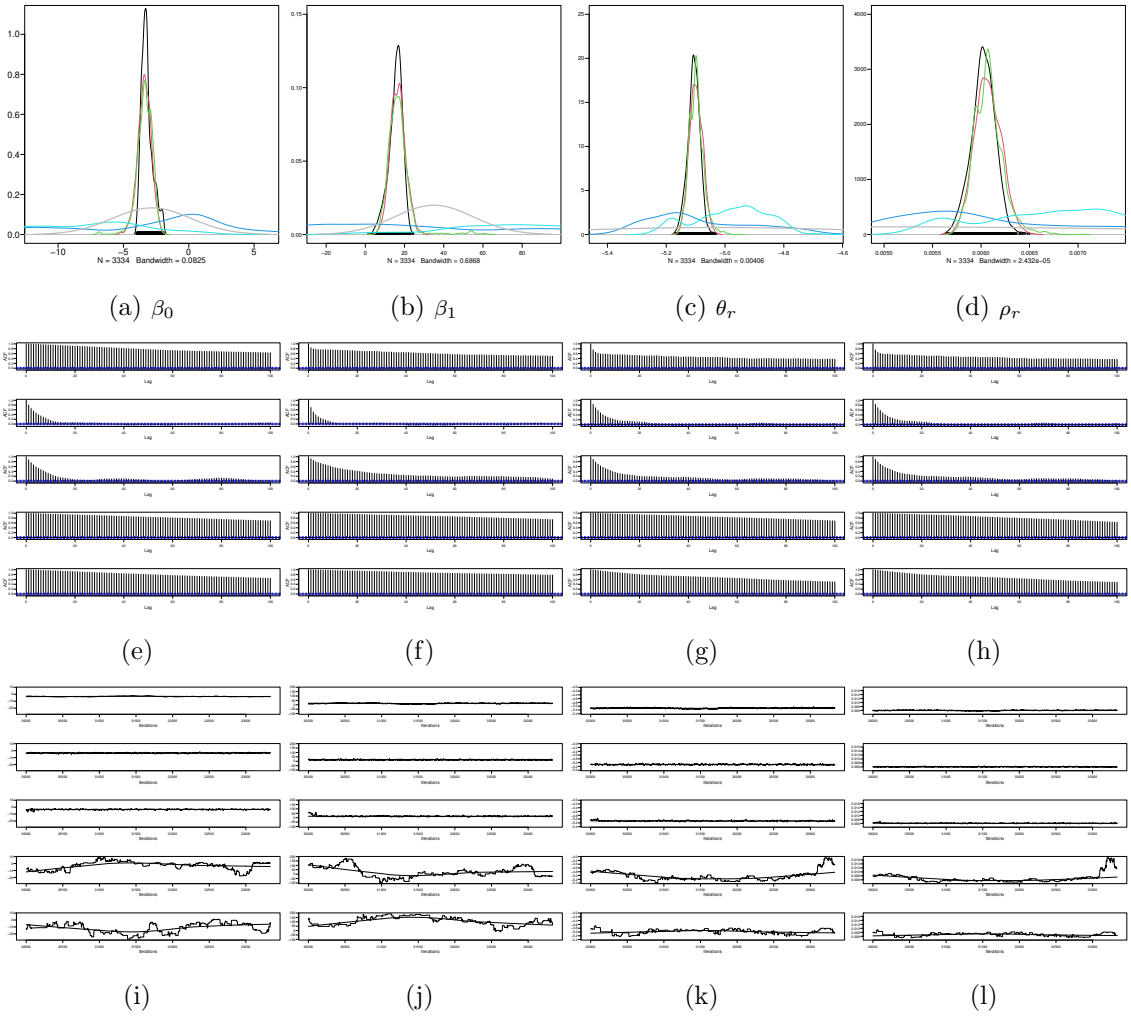


(a)



(b)

Figure 3.28: Top: Posteriors estimations (colored lines) from pMCMC with initial values: β_0 =-4.89, β_1 =35, and θ_r =-5.03 (first column β_0, second column β_1, third column θ_r, and fourth column ρ_r). Normal priors for the parameters β_0, β_1, θ_r (gray line). The colored lines for the posteriors correspond to the selected proposals, see Table 3.4. Middle: The estimated ACF of the Markov chain. Each row correspond to the proposal specified in Table 3.4. Bottom: The state of the Markov chain. Each row correspond to the proposal specified in Table 3.4. Thin parameter 20.

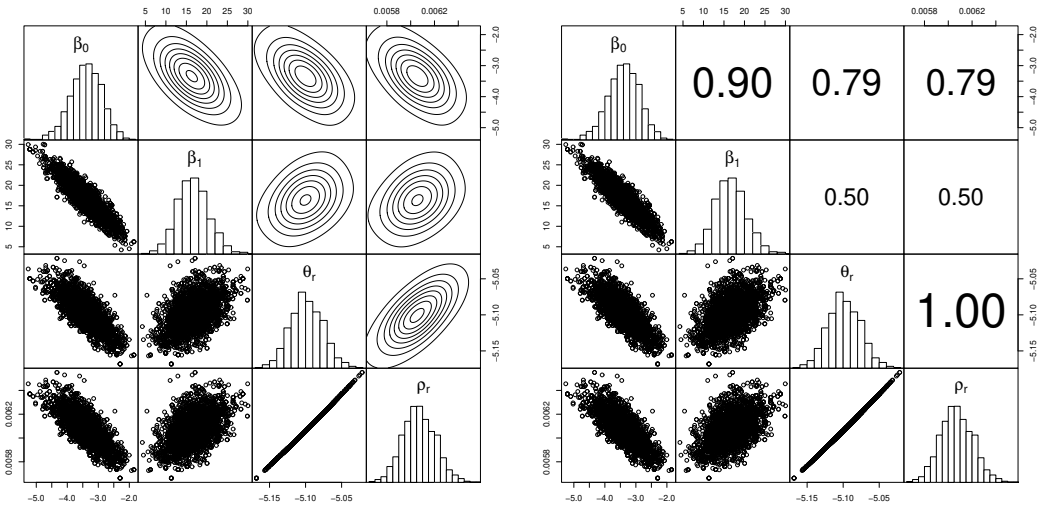


(a) β_0 (b) β_1 (c) θ_r (d) ρ_r

(e) (f) (g) (h)

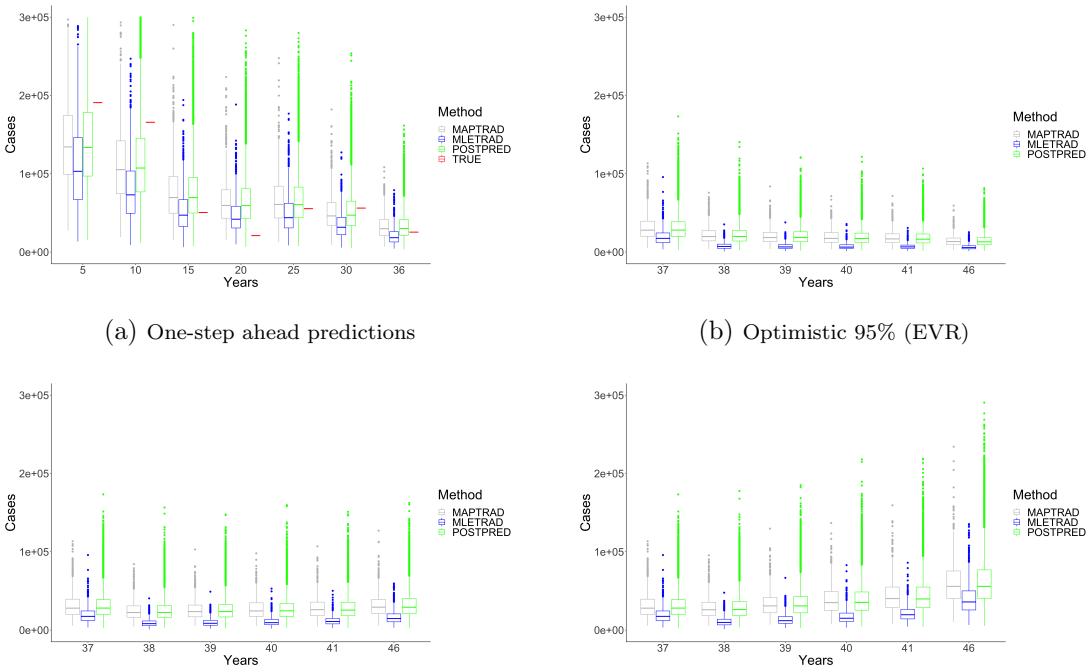
(i) (j) (k) (l)

Figure 3.29: Parameters posterior estimation relationship under proposal (red) (Table 3.4). The Pearson correlation between $\hat{\beta}_0$ and $\hat{\beta}_1$ is 0.90, between $\hat{\beta}_0$ and $\hat{\theta}_r$ is 0.79, and between $\hat{\beta}_1$ and $\hat{\theta}_r$ is 0.50.



(a) (b)

Figure 3.30: Comparison of the one-step ahead predictions and forecast using estimations from IF2 ($MLETRAD$), $MAPTRAD$ (taking the mean of the posterior distribution), and $POSTPRED$ (taking random draws from the posterior, respectively). Three scenarios of the effective vaccination rate (EVR) are evaluated optimistic (95%), moderate (75%), and pessimistic (50%).



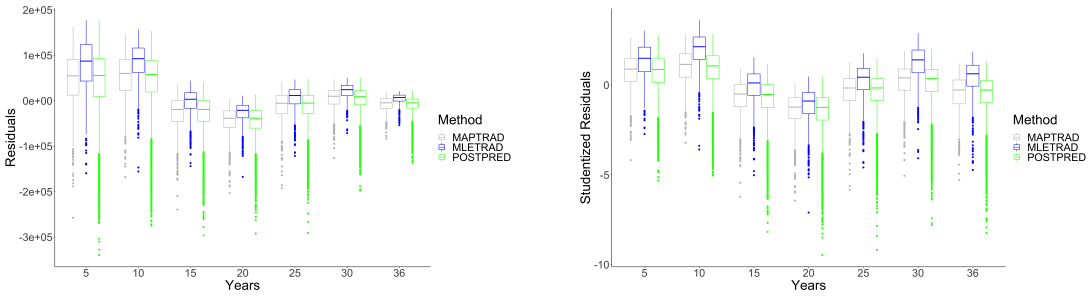
(a) One-step ahead predictions

(b) Optimistic 95% (EVR)

(c) Moderate 75% (EVR)

(d) Pessimistic 50% (EVR)

Figure 3.31: Residuals and studentized residuals one-step ahead predictions using the three methods $MLETRAD$, $MAPTRAD$, and $POSTPRED$. Three scenarios of the effective vaccination rate (EVR) are evaluated optimistic (95%), moderate (75%), and pessimistic (50%).

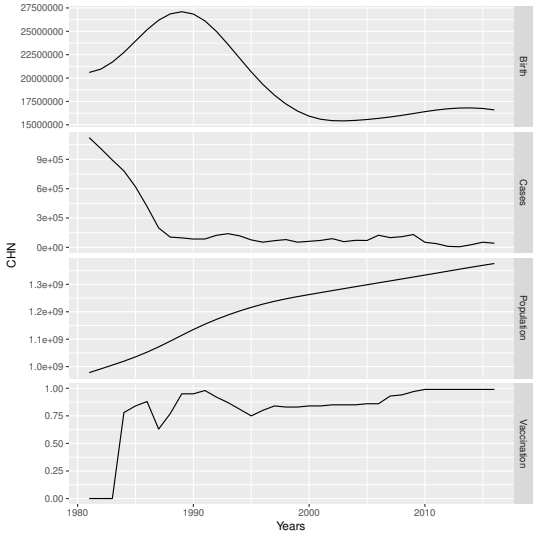


(a) Residuals

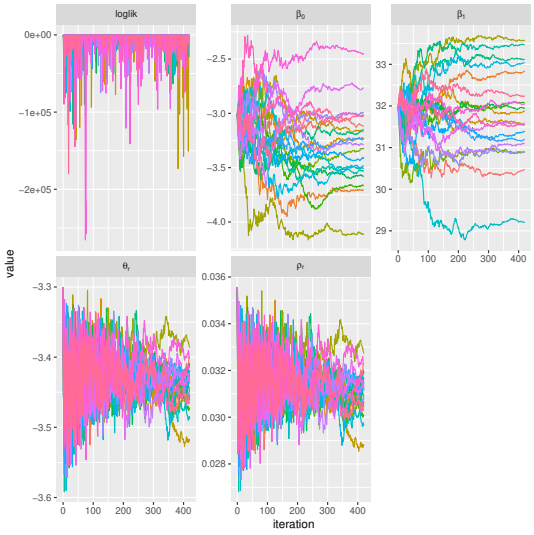
(b) Studentized residuals

China (CHN)

Figure 3.32: (a) Data cases from China refer to the number of births, number of reported measles cases, population size, and the effective vaccination rate. (b) Parameter estimation by using IF2 20 independent chains are evaluated, 420 iterations and a cooling parameter of 0.7

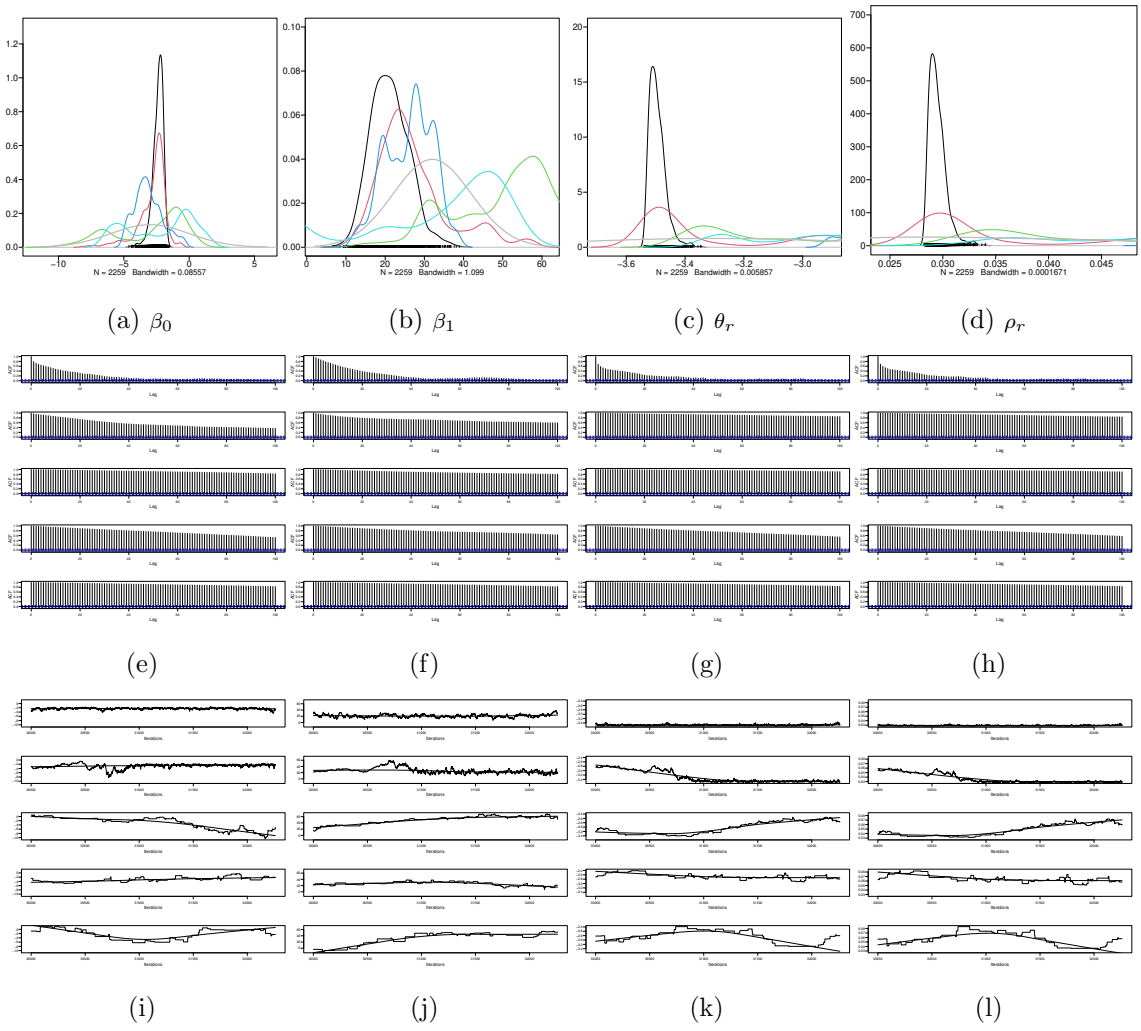


(a)



(b)

Figure 3.33: Top: Posteriors estimations (colored lines) from pMCMC with initial values: β_0 =-4.89, β_1 =35, and θ_r =-5.03 (first column β_0, second column β_1, third column θ_r, and fourth column ρ_r). Normal prior for the parameters β_0, β_1, θ_r (gray line). The colored lines for the posteriors correspond to the selected proposals, see Table 3.4. Middle: The estimated ACF of the Markov chain. Each row correspond to the proposal specified in Table 3.4. Bottom: The state of the Markov chain. Each row correspond to the proposal specified in Table 3.4. Thin parameter 30.

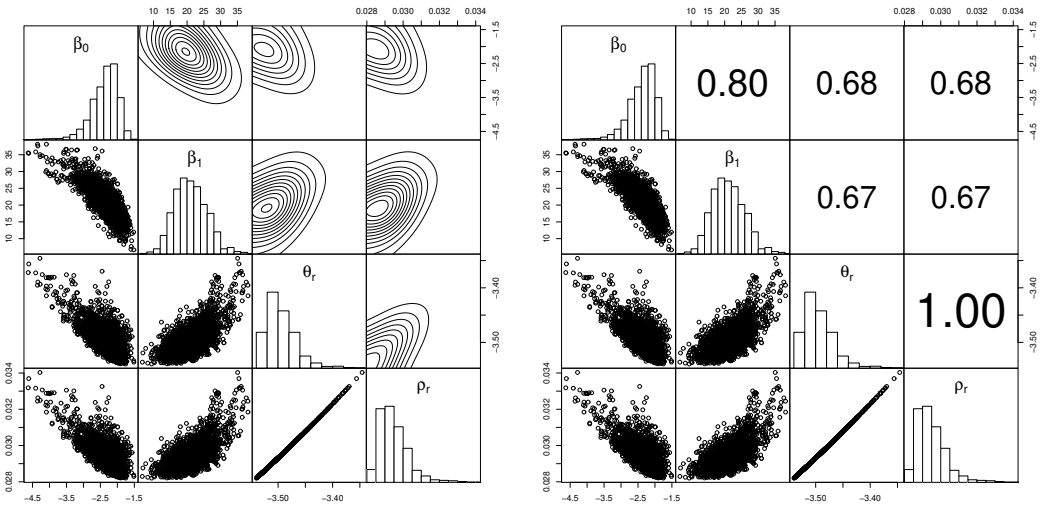


(a) β_0 (b) β_1 (c) θ_r (d) ρ_r

(e) (f) (g) (h)

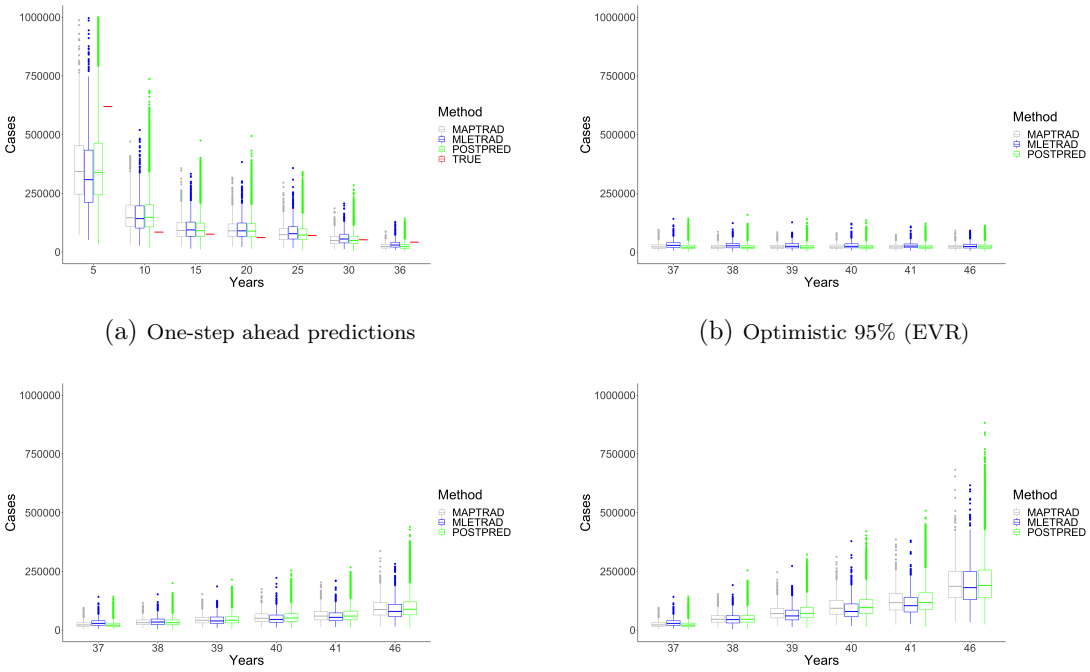
(i) (j) (k) (l)

Figure 3.34: Parameters posterior estimation relationship under proposal (black) (Table 3.4). The Pearson correlation between $\hat{\beta}_0$ and $\hat{\beta}_1$ is 0.80, between $\hat{\beta}_0$ and $\hat{\theta}_r$ is 0.68, and between $\hat{\beta}_1$ and $\hat{\theta}_r$ is 0.67.



(a) (b)

Figure 3.35: Comparison of the one-step ahead predictions and forecast using estimations from IF2 ($MLETRAD$), $MAPTRAD$ (taking the mean of the posterior distribution), and $POSTPRED$ (taking random draws from the posterior, respectively). Three scenarios of the effective vaccination rate (EVR) are evaluated optimistic (95%), moderate (75%), and pessimistic (50%).



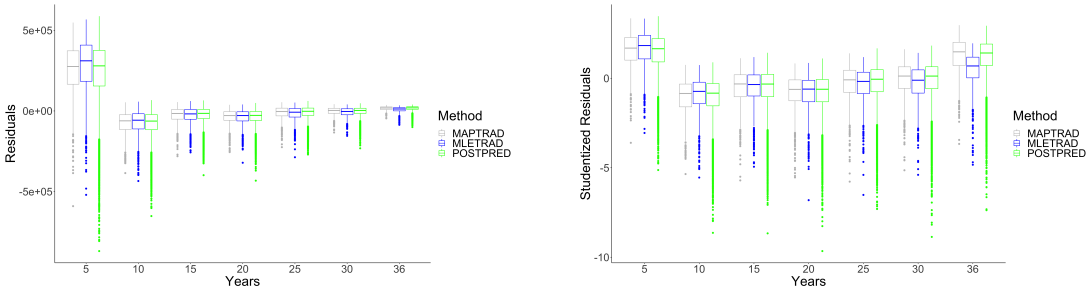
(a) One-step ahead predictions

(b) Optimistic 95% (EVR)

(c) Moderate 75% (EVR)

(d) Pessimistic 50% (EVR)

Figure 3.36: Residuals and studentized residuals one-step ahead predictions using the three methods $MLETRAD$, $MAPTRAD$, and $POSTPRED$. Three scenarios of the effective vaccination rate (EVR) are evaluated optimistic (95%), moderate (75%), and pessimistic (50%).

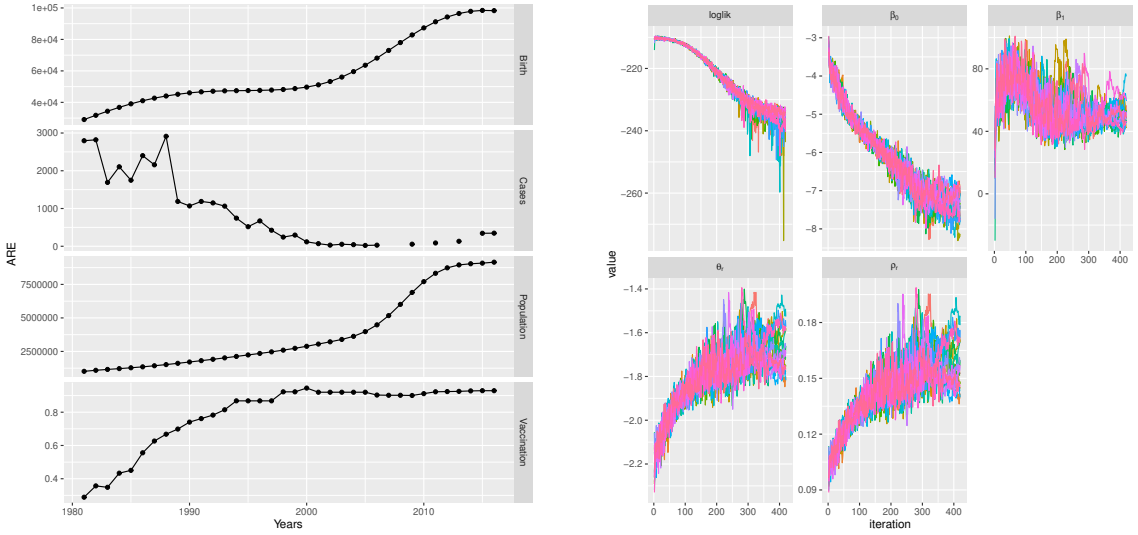


(a) Residuals

(b) Studentized residuals

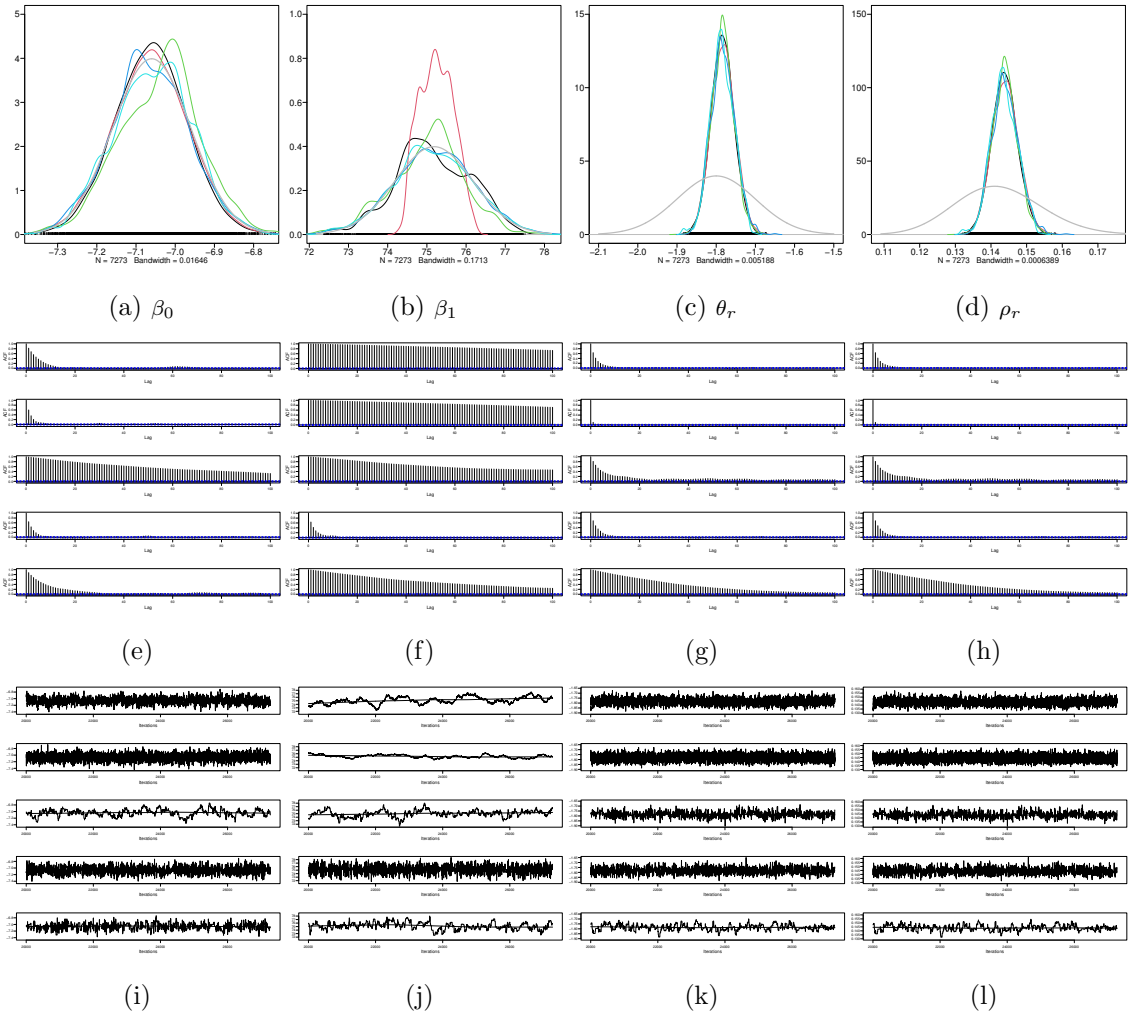
United Arab Emirates (ARE)

Figure 3.37: (a) Data cases from United Arab Emirates refer to the number of births, number of reported measles cases, population size, and the effective vaccination rate. (b) Parameter estimation by using IF2 20 independent chains are evaluated, 420 iterations and a cooling parameter of 0.7 (we omitted the two first iterations for a better visualization).



(a) (b)

Figure 3.38: Top: Posteriors estimations (colored lines) from pMCMC with initial values: β_0 =-4.89, β_1 =35, and θ_r =-5.03 (first column β_0, second column β_1, third column θ_r, and fourth column ρ_r). Normal prior for the parameters β_0, β_1, θ_r (gray line). The colored lines for the posteriors correspond to the selected proposals, see Table 3.4. Middle: The estimated ACF of the Markov chain. Each row correspond to the proposal specified in Table 3.4. Bottom: The state of the Markov chain. Each row correspond to the proposal specified in Table 3.4. Thin parameter 30.



(a) β_0 (b) β_1 (c) θ_r (d) ρ_r

(e) (f) (g) (h)

(i) (j) (k) (l)

Figure 3.39: Parameters posterior estimation relationship under proposal (black) (Table 3.4). The Pearson correlation between $\hat{\beta}_0$ and $\hat{\beta}_1$ is 0.80, between $\hat{\beta}_0$ and $\hat{\theta}_r$ is 0.68, and between $\hat{\beta}_1$ and $\hat{\theta}_r$ is 0.67.

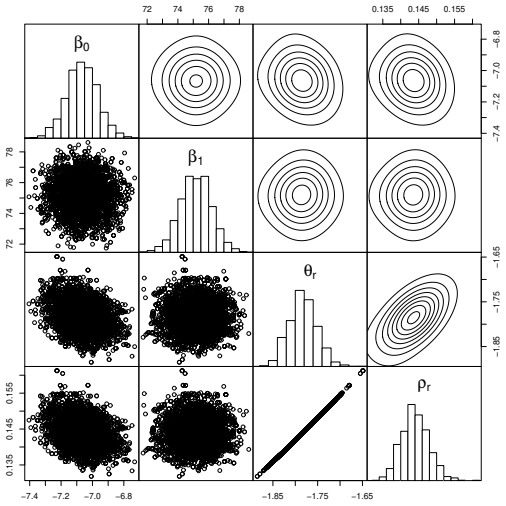
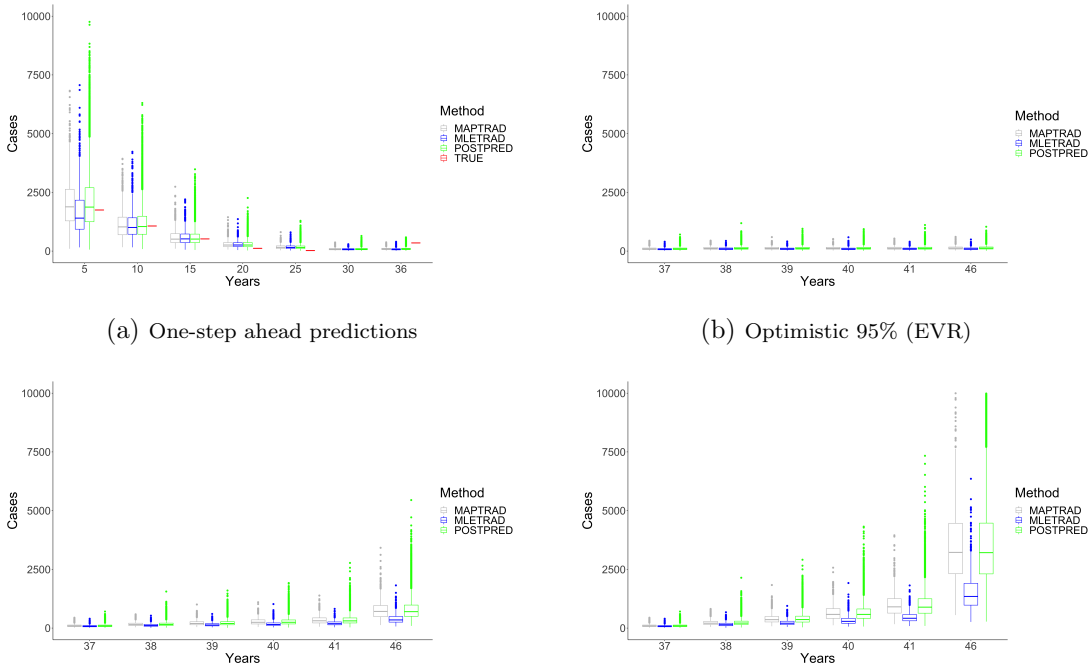


Figure 3.40: Comparison of the one-step ahead predictions and forecast using estimations from IF2 ($MLETRAD$), $MAPTRAD$ (taking the mean of the posterior distribution), and $POSTPRED$ (taking random draws from the posterior, respectively). Three scenarios of the effective vaccination rate (EVR) are evaluated optimistic (95%), moderate (75%), and pessimistic (50%).



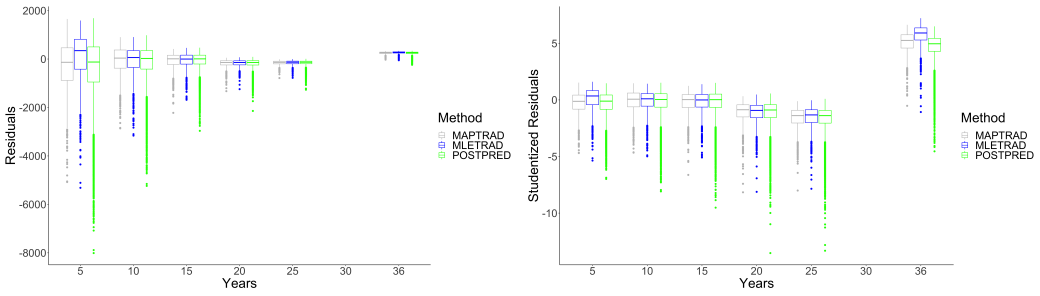
(a) One-step ahead predictions

(b) Optimistic 95% (EVR)

(c) Moderate 75% (EVR)

(d) Pessimistic 50% (EVR)

Figure 3.41: Residuals and studentized residuals one-step ahead predictions using the three methods $MLETRAD$, $MAPTRAD$, and $POSTPRED$. Three scenarios of the effective vaccination rate (EVR) are evaluated optimistic (95%), moderate (75%), and pessimistic (50%).



3.5 Conclusion

A Bayesian estimation that uses particle MCMC (pMCMC) was implemented in the context of the non-linear state space model proposed in Eilertson *et al.* (2019) without considering the age classes. Also, we implemented iteration filtering (IF2) as a supporting method. We assessed the proposed Bayesian estimation method using both simulation examples and real data sets. Although both the Bayesian method throughout pMCMC and IF2 exhibit similar results, the Bayesian approaches can produce credible intervals for the parameter estimations. Also, we evaluated the posterior estimates' robustness by considering different hyper-parameter specifications for the selected priors. Furthermore, choosing several proposals allowed us to improve the mixing rate.

Addressing simulation examples with three and four parameters for the model proposed in Eilertson *et al.* (2019) allowed us to evaluate the proposed Bayesian estimation procedure. In the case of the fourth parameter σ_e^2, we found that the estimates of the posterior densities are sensitive to the choice of priors. Regarding this fact, our results mimic the ones found by Eilertson *et al.* (2019). We also evaluated the model by performing one-step ahead predictions and forward predictions. The calculation of these predictions was based on three methods, one by using the estimations from IF2 ($MLETRAD$) and the other by using the results from pMCMC: a MAP estimations, taking the mode from the estimates of the posterior density ($MAPTRAD$) and posterior predictions ($POSPRED$). Moreover, we obtained one-step ahead predictions from four different priors whose results allowed us to confirm the robustness of the proposed Bayesian estimation. Besides considering the posterior estimates from the four priors, the forward predictions also evaluated three scenarios for the effective vaccination rate: optimistic, moderate, and pessimistic.

For the application to observed data, we proposed a Bayesian approach to fit the model proposed by Eilertson *et al.* (2019). We took advantage of the faster implementation of the IF2 algorithm to find reasonable initial conditions S_0, initial parameter values, and priors to perform the estimations using the pMCMC algorithm. Although we fitted the model using 12 countries' data sets, we reported results for three of them India, China, and United Arab Emirates. The countries showed similarities in the dynamic of the reported measles cases, a notorious decrease in the number of reported cases from 1990. The convergence of the parameter *probability of reporting a case*, θ_r, is common for the three countries according to the IF2 algorithm. However, the convergence for the parameters β_0 and β_1 is not apparent for China and India. The estimations from the two algorithms are similar for the three countries. However, the one-step ahead predictions from IF2 for India have a slightly better performance than the predictions from pMCMC. For the other two countries, the predictions with the method, posterior predictions, exhibit better performance.

Some weaknesses of the used algorithms were identified in countries with few reported cases. Notably, in the case of Colombia, the number of reported cases since 2005 is close to zero. These dynamics are typical for most South America countries. Also, the dynamics of the reported cases with extreme observations like Ethiopia or Congo, fitting the model is a difficult task. Specifically, in the case of Ethiopia, the extreme observations happened mainly before 1988, and subsequent removing this part of the sample, the estimation became problematic because of the scarcity of data.

Chapter 4

A BIRTH-DEATH MODEL WITH THRESHOLD MECHANISM

4.1 Abstract

A birth-death process is proposed to model the unobserved population size of a disease vector. This model considers the transmission effect to a second species by defining a stochastic process based on the vector population's size. This third element of the model measures a potential threshold event when the host species' population size surpasses a certain level yielding a higher transmission rate. A maximum likelihood procedure is developed for this model which combines particle filtering with the EM algorithm.

4.2 Introduction

A birth-death process is a special class of continuous-time Markov chain that is often used to study how the number of individuals in a population change over time. However, their applications are not restricted to this area Lanchier (2017). In a birth-death (BD) process, the state of the system is the number of individuals. Two reactions can occur: *births*, where the number of individuals increases by one, and *deaths*, where the number of individuals decreases by one. Instead of these considerations, when the rates depend on the population dynamic, this kind of modeling becomes more realistic in various applications. This kind of dependence is called a *general* BD process, see Crawford *et al.* (2014).

Statistical inference over general BD data observations depends on transition probabilities. Some authors have emphasized the calculation of the transitions probabili-

ties in terms of continued fractions expression for the Laplace transform of the transition probabilities, see for instance Crawford and Suchard (2012), Crawford *et al.* (2014). Nevertheless, Crawford *et al.* (2014) are pioneers in inference over the set of parameters in a general BD process using likelihood-based inference. Other approaches to the estimation process consider the case of likelihood-free alternatives such as approximate Bayesian computation, see Drovandi and Pettitt (2011) or Owen *et al.* (2014).

Since a general BD process tracks only one population, it limits some biological or epidemiological applications. In many situations, the interaction between two or more species becomes crucial to understanding the dynamics of these populations, such as competition or infection events. Extensions of the BD process to bivariate processes are recent topics in the literature, such as the competition process, which allows birth and death events as well as individual movement events from one population to the other, see Reuter (1961), and Ho *et al.* (2018). Further extension is the branching processes to study bivariate populations in the context of birth-death-shift process Xu *et al.* (2014). This extension allows multiple events to occur per observation interval. A third extension which is a subclass of a competition process, is proposed by Ho *et al.* (2018). They develop an efficient method to compute the transition probabilities in this context.

This chapter proposes an extension of the general BD process to evaluate the interaction between two populations. We feature a model that studies the effect of a disease vector population size on a second affected population. We assume that the vector population follows a birth-death process with a threshold mechanism (BDTM) as it affects the second population as we present in section 4.4. We also assume that the vector population is unobserved. Furthermore, the second population follows a non-homogenous Poisson process when condition on the vector process with

a transition rate given by the dynamic of the vector population. Our motivation for this paper is to estimate the threshold of the vector population size that causes an increasing transmission to the second population.

Since the estimation of transition probabilities for this kind of process is a difficult task Ho *et al.* (2018), to the estimation process, we propose a method based on the EM algorithm Dempster *et al.* (1977) and Sequential Monte Carlo (SMC) approach Doucet *et al.* (2010). In section 4.4 we derive likelihood-based closed-form expressions for the estimators. These closed form expressions depend on the realization paths of the unobserved vector population, which is a continuous-time Markov process, and the observed data of the population size from the affected group (number of infected), which is a set of discrete-time observations. In this sense, using an SMC method, we simulate exact continuous paths inside each time interval for the unobserved vector population, corresponding to the discrete-time observations. Thus, based on these simulated paths and the observed data, a set of weighted paths approximates the unobserved vector population size.

4.3 Threshold Mechanism

Before presenting a full model, we will give some literature review related to the model we will discuss in the next section. A Poisson process with change points can be defined as either changing over states or over time. Changes points over the state and over time are related to nonlinearity and non-stationarity, respectively, see Tong (2015). These kinds of models, mainly in the econometrics literature, are also known as models with structural breaks, see Aue and Horváth (2013). In real applications, discrimination between these two models is a difficult task, mainly because we have just one realization, which is why the detection of changing points is relevant Tong (2015).

From the perspective of change points over time Akman and Raftery (1986); Raftery and Akman (1986) propose a test to identify change points in a Poisson process by following an asymptotic and Bayesian approach, respectively. They assume that the Poisson process has rate λ_1 for the period of time $0 \leqslant t \leqslant \tau$ before the change (*jump*), and the poisson process has rate λ_2 after the jump in the time interval $\tau \leqslant t \leqslant T$. Also these authors propose an interval estimate for the rates λ_1, λ_2, and the time τ where the change happens. Loader (1992) proposes a log-linear model with change point to the rate function of a non-homogenous Poisson process as long as a model selection procedure. Thus, the rate function $\lambda(t)$ is fitted with an exponential function $\lambda(t) = \exp\{a+bt\}$ if $0 \leqslant t \leqslant \tau$, and $\lambda(t) = \exp\{a+\delta+bt\}$ if $\tau \leqslant t \leqslant T$ where a, b, δ, and τ are unknown parameters (τ denotes the time for the change point). West and Ogden (1997), a maximum likelihood point estimate and Bayesian-based interval estimator are obtained for a change point in a Poisson process. Given the discontinuity in the changes of the rate of the Poisson process "jumps" to finding the maximum likelihood estimators for the three parameters, the rate before the jump (λ_1), the rate after the jump (λ_2), and the jump parameter τ a range over a set of points on the interval $[0, T]$ is considered for τ. Thus, the likelihood, $\hat{\lambda}_1(\tau)$ and $\hat{\lambda}_2(\tau)$ are computed for each valor of τ. For the interval estimation, something similar is developed by defining a prior for the change-point τ. Boudjellaba *et al.* (2001) focuses on exact distributions of test statistics to produce confidence intervals instead of asymptotic approximations.

From the perspective of change points over the state, Wang *et al.* (2014) frames and analyzes a self-excited threshold Poisson autoregression process. This model assumes that the observations $\{Y_t\}$ follow a Poisson distribution conditioned on an intensity process that is defined with two regimes. More specifically, the intensity process $\{\lambda_t\}$ follows a GARCH model defined by Bollerslev (1986) to fit the conditional

heteroskedasticity by using the fact that the Poisson mean coincides with its variance. Regarding the estimation of the threshold parameter, they consider searching over all candidates of values on a given interval $[0, r^*]$. However, since the number of values in this range could be large and if the bound is too broad, both cases have limitations for the estimation process because the former depends on computational resources, which could be limited. The last number of observations might not be enough to guarantee consistent estimates. To solve these inconveniences, they reduce the search interval by considering a range for a given empirical quantile from the observations $\{Y_t\}$.

Other methodologies to detect and analyze change points over state are by following the mixture distributions approach. For example, Smith (1989); Davison and Smith (1990) propose statistical models to fit exceedances over high thresholds. The simplest statistical models considered by these authors to fit exceedances over a given threshold are based on the Poisson point process. More complicated models are defined based on generalized Pareto distributions. This kind of representation is often referred to as Poisson-GPD Scarrott and MacDonald (2012). Under the generalized Pareto distributions approach Behrens *et al.* (2004) analyze extremal events considering explicitly the uncertainty about the threshold where it is directly estimated. To estimate the threshold, they propose both a parametric form and a generalized Pareto distribution to fit the observations below and beyond the threshold, respectively. The estimation routine uses Bayesian methods. Subsequently, do Nascimento *et al.* (2012) propose a semiparametric approach for extreme value density estimation. For observations under a given threshold, a combination of a mixture of Gamma distributions is used; while for the observations above the threshold, the generalized Pareto distribution is adopted. This methodology allows for the estimation of all model parameters, even the threshold parameter. In addition, future predictions are

provided under this approach. Other papers that used the mixture distributions approach include Frigessi *et al.* (2002); Tancredi *et al.* (2006); Diebolt *et al.* (2005), among others. Now, we turn to the definition of the model.

4.4 Model

Now we define the model that we will use. This model has two coupled equations: the first X_t measures a population size using a birth-death process, the second corresponds to the observation model and includes a threshold mechanism.

$$X_t = X_0 + Y_1\bigg(\int_0^t (\alpha_1 X_s + a)ds\bigg) - Y_2\bigg(\int_0^t (\alpha_2 X_s + b) I_{\{X_s>0\}} ds\bigg), \tag{4.1}$$

$$\begin{aligned} g(x_t) &= \theta_1 x_t + \theta_2 x_t I_{\{X_t>c\}}, \\ Z_t &= Y_3\bigg(\int_0^t g\big(X_s\big)ds\bigg), \end{aligned} \tag{4.2}$$

where $I_{\{X_t>c\}} = 1$ if $X_t > c$ and zero otherwise; and

$$W_n = Z_{t_n} - Z_{t_{n-1}}. \tag{4.3}$$

The model in (4.1) is a general birth-death process (GBD) in a *continuous time* with immigration and emigration rates (Crawford *et al.* (2014)). We propose this model to understand the dynamic of a vector population with a threshold mechanism. Thus, X_t measures the population size (x) of the disease vector at times $t \geqslant 0$. In this sense, from state $X_t = x$ transitions to state $x + 1$ take place with exponential rate $\lambda_x = \alpha_1 x + a$ where α_1 is the birth rate and a the immigration rate. Likewise, exponential transitions to state $x - 1$ take place with rate $\mu_x = \alpha_2 x + b$ where α_2 is the death rate and b the emigration rate. In addition, the model in (4.2) represents a threshold effect, *infection transmissions* from vector population to a

second species. In 4.18, Z_{t_n} counts the number of infecteds, in the second species, that become up to time t_n; t_n in the continuous time framework is the n observation time. Furthermore, $g(x_t) = (\theta_1 + \theta_2 I_{\{X_t > c\}})x_t$ is the infection rate where θ_1 denotes the number of infections per vector, θ_2 denotes the additional infection rate per vector when the population size of the vectors is over the threshold, and c is the threshold parameter.

4.5 Inference for the Model

Since the vector population is often not fully observed, in the proposed model, we assume that the process $\{X_t\}$ is unobserved. Instead, we assume tracking the reports of the number of infected. These reports are available in most cases in discrete-time, monthly or annually for instance. Under these conditions, the estimation process is nontrivial. To fit the proposed model, we need to predict the states of $\{X_t\}$, which have a discrete state space, and to estimate the parameters of the model. For the third element of the model, we will create approximation of the model to ensure that the likelihood function is continuous and differentiable in c. Thus, instead of using a strict threshold, we will optimize over this approximation, and that will be easier if we use a smooth function. So, we replace the indicator in (4.2) with a logistic function $\Phi_c(x_t) = \exp\{((x_t - c)/\sigma)\}$ which also has a tuning parameter σ. As a result, when the tuning parameter is small, this approximates the threshold model.

Crawford *et al.* (2014) develop a methodology to estimate the transition probabilities for a given GBD using Laplace transforms (Crawford and Suchard (2012)). Also, these authors propose a modification of the EM algorithm Dempster *et al.* (1977) by defining a surrogate function for the (M-step) following the minorize-maximization (MM) (Lange (2012)) to estimate the birth and death parameters under several conditions. The notation used by Crawford *et al.* (2014) to establish the log-likelihood

function includes variables U_k, D_k, and T_k that accumulate the number of births, deaths and the time spend in each state k. Thus, U_k, D_k represent the number of *up* steps and *down* steps for the GBD $\{X_t\}$ up to time t. Thus, the log-likelihood for a continuous observed process has a simple form where the sum is over all possible states k:

$$l(\theta) = \sum_{k=0}^{\infty} \Big[U_k \log(\lambda_k(\theta)) + D_k \log(\mu_k(\theta)) - (\lambda_k(\theta) + \mu_k(\theta)) T_k \Big] \tag{4.4}$$

where λ_k and μ_k are the birth and death rates, respectively. However, when the process $\{X_t\}$ is observed in a discrete time, the estimation of the transition rates is difficult because the state path between observations is unobserved. Thus, the quantities U_k, D_k, and T_k are unknown for every state k. Since the maximization of the log-likelihood is not possible due to the missing observations between the discrete observations, the MM algorithm is implemented to impute the missing information between time points. Thus, the MM algorithm allows to do inference over the parameters.

In the context of our proposed model, we cannot apply these methods directly because we do not observe the GBD $\{X_t\}$.

4.5.1 Log-likelihood for the Proposed Model

To specify the likelihood function, we follow the expression in (4.4), but we define a new notation. The likelihood function in (4.5) can be read as follows: the last term comes from the remainder of time in the Markov chain after the last event, the second to the last term comes from Poisson observations conditioned on X_s for $s \in [t_{n-1}, t_n]$ and t_n in the continuous time framework is the n observation time. The first and second terms are counting the transitions between states of the continuous Markov chain, so those count the number of *up* steps and *down* steps along to the

corresponding rates

$$
\begin{aligned}
L = \prod_{n=1}^{N} & \exp\left\{ \int_{t_{n-1}}^{t_n} \log(\alpha_1 X_s + a) dU_{1s} \right\} \exp\left\{ \int_{t_{n-1}}^{t_n} \log(\alpha_2 X_s + b) dU_{2s} \right\} \\
& \exp\left\{ (Z_{t_n} - Z_{t_{n-1}}) \log\left(\int_{t_{n-1}}^{t_n} g\big(X_s\big) ds \right) - \log(Z_{t_n} - Z_{t_{n-1}})! \right\} \\
& \times \exp\left\{ - \int_{t_{n-1}}^{t_n} \Big[a + b + (\alpha_1 + \alpha_2) X_s \Big] ds \right\} \\
& \times \exp\left\{ - \int_{t_{n-1}}^{t_n} \Big[\theta_1 + \theta_2 \Phi_c(X_s) \Big] X_s ds \right\},
\end{aligned}
\tag{4.5}
$$

where

$$
\begin{aligned}
U_1(t) &= Y_1\bigg(\int_0^t (\alpha_1 X_s + a) ds \bigg), \text{and} \\
U_2(t) &= Y_2\bigg(\int_0^t (\alpha_2 X_s + b) I_{\{X_s > 0\}} ds \bigg).
\end{aligned}
$$

Given a set of observations $\{Z_{t_1}, Z_{t_2}, ..., Z_{t_n}\}$, the log-likelihood function for n time intervals is

$$
\begin{aligned}
l = \log(L) = \sum_{n=1}^{N} \bigg(& \int_{t_{n-1}}^{t_n} \log(\alpha_1 X_s + a) dU_{1s} + \int_{t_{n-1}}^{t_n} log(\alpha_2 X_s + b) dU_{2s} \\
& + (Z_{t_n} - Z_{t_{n-1}}) \log\left(\int_{t_{n-1}}^{t_n} g\big(X_s\big) ds \right) - \log(Z_{t_n} - Z_{t_{n-1}})! \\
& - \int_{t_{n-1}}^{t_n} \Big[a + b + (\alpha_1 + \alpha_2) X_s \Big] ds \\
& - \int_{t_{n-1}}^{t_n} \Big[\theta_1 + \theta_2 \Phi_c(X_s) \Big] X_s ds \bigg),
\end{aligned}
\tag{4.6}
$$

We switched notation because it allows us to write what happens in each interval (given by the discrete observations) as an integral, which helps us when we

think about the particle filtering and when we think about changing the number of states. To maximize the log-likelihood function we apply the minorize-maximize (MM) algorithm, Lange (2012). To do this we minorize the log-likelihood function $l(\Theta)$ in (4.6) by finding a surrogate function (M) such that $l(\Theta) \geq M(\Theta|\Theta^{(m)})$ and $l(\Theta^{(m)}) = M(\Theta^{(m)}|\Theta^{(m)})$, where Θ is the vector of parameters. The main difficulty to maximize (4.6) is the impossibility to separate the terms inside the $\log(\cdot)$ function. Therefore the minorization step is applied for these $\log(\cdot)$ terms. First, we use the fact that the $\log(\cdot)$ terms are log-concave. Thus

$$
\begin{aligned}
\log\left(\alpha_1 X_s + a\right) &= \log\left(\rho\frac{\alpha_1 X_s}{\rho} + (1-\rho)\frac{a}{(1-\rho)}\right), \text{ with } \rho \in [0,1] \\
&\geqslant \rho\log\left(\frac{\alpha_1 X_s}{\rho}\right) + (1-\rho)\log\left(\frac{a}{1-\rho}\right), \text{by log-concave property} \\
&\geqslant \rho\log\left(\frac{\alpha_1 X_s}{\rho}\right) + (1-\rho)\log\left(\frac{a}{1-\rho}\right) + \rho\log\left(\rho^2\right) + \\
&\quad (1-\rho)\log\left((1-\rho)^2\right), \\
&\quad (\text{since } 0 \leqslant \rho \leqslant 1, \log\left(\rho^2\right) < 0, \log\left((1-\rho)^2\right) < 0) \\
&= \rho\log\left(\rho\alpha_1 X_s\right) + (1-\rho)\log((1-\rho)a) \\
&= \rho\log\left(\rho\alpha_1\right) + \rho\log\left(X_s\right) + (1-\rho)\log((1-\rho)a)
\end{aligned}
$$

Now, we choose

$$
\rho = \rho_1^{(m)} = \frac{\alpha_1^{(m)} X_s}{\alpha_1^{(m)} X_s + a^{(m)}} \tag{4.7}
$$

for a given $\alpha_1^{(m)}$ and $a^{(m)}$. Similarly, we have for the term

$$\log\left(\alpha_2 X_s + b\right) \geqslant \rho_2 \log\left(\rho_2 \alpha_2\right) + \left(1-\rho_2\right) \log\left(\left(1-\rho_2\right) b\right)$$

with

$$\rho_2 = \rho_2^{(m)} = \frac{\alpha_2^{(m)} X_s}{\alpha_2^{(m)} X_s + b} \tag{4.8}$$

for a given $\alpha_2^{(m)}$ and $b^{(m)}$. For the term

$$\begin{aligned}
\log\left(\int_{t_{n-1}}^{t_n} g\left(X_s\right) ds\right) &= \log\left(\left(t_n - t_{n-1}\right) \int_{t_{n-1}}^{t_n} g\left(X_s\right) \frac{1}{t_n - t_{n-1}} ds\right) \\
&\geqslant \log\left(t_n - t_{n-1}\right) + \int_{t_{n-1}}^{t_n} \log\left(g\left(X_s\right) \frac{1}{t_n - t_{n-1}}\right) ds, \\
&\quad \text{(Jensen's inequality)}, \\
&= \log\left(t_n - t_{n-1}\right) - \left(t_n - t_{n-1}\right) \log\left(t_n - t_{n-1}\right) + \\
&\quad \int_{t_{n-1}}^{t_n} \log\left(g\left(X_s\right)\right) ds.
\end{aligned}$$

Since $g\left(X_s\right) = \left(\theta_1 + \theta_2 \Phi_c\left(X_s\right)\right) X_s$, by minorization step we have

$$\begin{aligned}
\log\left(g\left(X_s\right)\right) &= \log\left[\left(\theta_1 + \theta_2 \Phi_c\left(X_s\right)\right) X_s\right] = \log\left(\theta_1 + \theta_2 \Phi_c\left(X_s\right)\right) + \log\left(X_s\right) \\
&\geq \rho_3 \log\left(\rho_3 \theta_1\right) + \left(1-\rho_3\right) \log\left(\left(1-\rho_3\right) \theta_2\right) + \left(1-\rho_3\right) \log\left(\Phi_c\left(X_s\right)\right) + \\
&\quad \log\left(X_s\right)
\end{aligned}$$

with

$$\rho_3 = \rho_3^{(m)} = \frac{\theta_1^{(m)}}{\theta_1^{(m)} + \theta_2^{(m)} \Phi_{c^{(m)}}\left(X_s\right)}. \tag{4.9}$$

By combining everything we have

$$
\begin{aligned}
l(\Theta) &\geq M(\Theta|\Theta^{(m)}) \\
&= \sum_{n=1}^{N} \Bigg(\int_{t_{n-1}}^{t_n} \left[\rho_1^{(m)} \log\left(\rho_1^{(m)} \alpha_1 \right) + \left(1 - \rho_1^{(m)} \right) \log\left(\left(1 - \rho_1^{(m)} \right) a \right) \right] dU_1(s) + \\
&\qquad \int_{t_{n-1}}^{t_n} \left[\rho_2^{(m)} \log\left(\rho_2^{(m)} \alpha_2 \right) + \left(1 - \rho_2^{(m)} \right) \log\left(\left(1 - \rho_2^{(m)} \right) b \right) \right] dU_2(s) + \\
&\qquad (z_n - z_{n-1}) \Bigg[\tau + \int_{t_{n-1}}^{t_n} \Big\{ \rho_3^{(m)} \log\left(\rho_3^{(m)} \theta_1 \right) + \\
&\qquad\qquad \left(1 - \rho_3^{(m)} \right) \log\left(\left(1 - \rho_3^{(m)} \right) \theta_2 \right) + \\
&\qquad\qquad \left(1 - \rho_3^{(m)} \right) \log\left(\Phi\left(X_s \right) \right) \Big\} ds \Bigg] - \log\left(z_n - z_{n-1} \right)! + \\
&\qquad \int_{t_{n-1}}^{t_n} \rho_1^{(m)} \log\left(X_s \right) dU_1(s) + \int_{t_{n-1}}^{t_n} \rho_2^{(m)} \log\left(X_s \right) dU_2(s) + \\
&\qquad\qquad (z_n - z_{n-1}) \int_{t_{n-1}}^{t_n} \log\left(X_s \right) ds \\
&\qquad - \int_{t_{n-1}}^{t_n} \Big[a + b + \left(\alpha_1 + \alpha_2 \right) X_s \Big] ds \\
&\qquad - \int_{t_{n-1}}^{t_n} \Big[\theta_1 + \theta_2 \Phi_c(X_s) \Big] X_s ds \Bigg),
\end{aligned}
$$

where $\tau = \log\left(t_n - t_{n-1} \right) \left(1 - \left(t_n - t_{n-1} \right) \right)$. Now letting

$$
Q\left(\Theta|\Theta^{(m)} \right) = E_{\Theta^{(m)}} \left(M\left(\Theta \mid \Theta^{(m)} \right) \mid Z_1, \cdots, Z_N \right) \tag{4.10}
$$

We approximate the calculation of this expectation by particle filtering. For each pair of adjacent time points, we simulate J *interval time trajectories* for $\{X_t\}$, that is, $X^{(j)}_{(t_{n-1},t_n]}$ with $n = 1, ..., T$ and $j = 1, ..., J$. Thus, the approximation of the expected value in (4.10) is given by

$$
Q\left(\Theta|\Theta^{(m)} \right) \approx \sum_{n=1}^{N} M\left(\Theta \mid \Theta^{(m)} \right) \omega_n^{(j)}, \tag{4.11}
$$

where the weights $\omega_n^{(j)}$ have a Poisson form to evaluate at $W_n = Z_n - Z_{n-1}$ with mean

$\int_{t_{n-1}}^{t_n} g(X_s^{(j)})ds$. To simplify the notation we refers to the weights as $\omega^{(j)}$ although they are calculated for each time interval. Thus,

$$
\begin{aligned}
&Q(\Theta|\Theta^{(m)}) \\
&= \sum_{n=1}^{N}\sum_{j=1}^{J}\left(\int_{t_{n-1}}^{t_n}\left[\rho_1^{(m,j)}\log\left(\rho_1^{(m,j)}\alpha_1\right)+\left(1-\rho_1^{(m,j)}\right)\log\left(\left(1-\rho_1^{(m,j)}\right)a\right)\right]dU_1^{(j)}(s)+\right. \\
&\qquad \int_{t_{n-1}}^{t_n}\left[\rho_2^{(m,j)}\log\left(\rho_2^{(m,j)}\alpha_2\right)+\left(1-\rho_2^{(m,j)}\right)\log\left(\left(1-\rho_2^{(m,j)}\right)b\right)\right]dU_2^{(j)}(s)+ \\
&\qquad (Z_n-Z_{n-1})\left[\tau+\int_{t_{n-1}}^{t_n}\left\{\rho_3^{(m,j)}\log\left(\rho_3^{(m,j)}\theta_1\right)+\right.\right. \\
&\qquad\qquad \left(1-\rho_3^{(m,j)}\right)\log\left(\left(1-\rho_3^{(m,j)}\right)\theta_2\right)+ \\
&\qquad\qquad \left.\left.\left(1-\rho_3^{(m,j)}\right)\log\left(\Phi\left(X_s^{(j)}\right)\right)\right\}ds\right]-\log\left(Z_n-Z_{n-1}\right)!+ \\
&\qquad \int_{t_{n-1}}^{t_n}\rho_1^{(m,j)}\log\left(X_s\right)dU_1^{(j)}(s)+\int_{t_{n-1}}^{t_n}\rho_2^{(m,j)}\log\left(X_s^{(j)}\right)dU_2^{(j)}(s)+ \\
&\qquad (Z_n-Z_{n-1})\int_{t_{n-1}}^{t_n}\log\left(X_s^{(j)}\right)ds \\
&\qquad -\int_{t_{n-1}}^{t_n}\left[a+b+\left(\alpha_1+\alpha_2\right)X_s^{(j)}\right]ds \\
&\qquad \left.-\int_{t_{n-1}}^{t_n}\left[\theta_1+\theta_2\Phi_c(X_s^{(j)})\right]X_s^{(j)}ds\right)\omega^{(j)},
\end{aligned}
\tag{4.12}
$$

where $\rho_1^{(m,j)}, \rho_2^{(m,j)}$, and $\rho_3^{(m,j)}$ are given in (4.7), (4.8), and (4.9) evaluated at $X_s^{(j)}$. Now from (4.12) we maximize, thus we obtain the estimators for each parameter, respectively as follow

$$
\frac{\partial Q}{\partial \alpha_1}=\sum_{n=1}^{N}\sum_{j=1}^{J}\omega^{(j)}\left(\frac{1}{\alpha_1}\int_{t_{n-1}}^{t_n}\rho_1^{(m,j)}dU_{1s}^{(j)}-\int_{t_{n-1}}^{t_n}X_s^{(j)}ds\right)=0,
$$

then

$$
\hat{\alpha}_1^{(m+1)}=\frac{\sum_{n=1}^{N}\sum_{j=1}^{J}w^{(j)}\int_{t_{n-1}}^{t_n}\rho_1^{(m,j)}dU_1^{(j)}(s)}{\sum_{n=1}^{N}\sum_{j=1}^{J}\omega^{(j)}\int_{t_{n-1}}^{t_n}X_s^{(j)}ds},
$$

$$\hat{a}^{(m+1)} = \frac{\sum_{n=1}^{N}\sum_{j=1}^{J} w^{(j)} \int_{t_{n-1}}^{t_n}(1-\rho_1^{(m,j)})dU_1^{(j)}(s)}{N},$$

$$\hat{\alpha}_2^{(m+1)} = \frac{\sum_{n=1}^{N}\sum_{j=1}^{J} w^{(j)} \int_{t_{n-1}}^{t_n}\rho_2^{(m,j)}dU_2^{(j)}(s)}{\sum_{n=1}^{N}\sum_{j=1}^{J}\omega^{(j)}\int_{t_{n-1}}^{t_n}X_s^{(j)}ds},$$

$$\hat{b}^{(m+1)} = \frac{\sum_{n=1}^{N}\sum_{j=1}^{J} w^{(j)} \int_{t_{n-1}}^{t_n}(1-\rho_2^{(m,j)})dU_2^{(j)}(s)}{N},$$

$$\hat{\theta}_1^{(m+1)} = \frac{\sum_{n=1}^{N}(Z_n - Z_{n-1})\sum_{j=1}^{J} w^{(j)} \int_{t_{n-1}}^{t_n}\rho_3^{(m,j)}ds}{\sum_{n=1}^{N}\sum_{j=1}^{J} w^{(j)} \int_{t_{n-1}}^{t_n}X_s^{(j)}ds}$$

$$\hat{\theta}_2^{(m+1)} = \frac{\sum_{n=1}^{N}(Z_n - Z_{n-1})\sum_{j=1}^{J} w^{(j)} \int_{t_{n-1}}^{t_n}\left(1-\rho_3^{(m,j)}\right)ds}{\sum_{n=1}^{N}\sum_{j=1}^{J} w^{(j)} \int_{t_{n-1}}^{t_n}\Phi_c\left(X_s^{(j)}\right)X_s^{(j)}ds}$$

For parameter c we first calculate the derivative of the function $\Phi_c(x) = \frac{e^{(x-c)/\sigma}}{1+e^{(x-c)/\sigma}}$ respect to c, this is:

$$\begin{aligned}\Phi_c'(x) = \frac{\partial\Phi_c(x)}{\partial c} =& \frac{\frac{-1}{\sigma}e^{(x-c)/\sigma}}{1+e^{(x-c)/\sigma}} + \frac{\frac{1}{\sigma}e^{(x-c)/\sigma}}{(1+e^{(x-c)\sigma})^2} \\ =& \frac{1}{\sigma}\Phi_c(x)\left(\frac{e^{(x-c)/\sigma}}{1+e^{(x-c)/\sigma}} - 1\right) \\ =& -\Phi_c(x)\frac{1/\sigma}{1+e^{(x-c)/\sigma}}.\end{aligned}$$

Hence, for the parameter c we have

$$\begin{aligned}\frac{\partial Q}{\partial c} =& \\ &\sum_{n=1}^{N}(Z_n - Z_{n-1})\sum_{j=1}^{J}\omega_j\int_{t_{n-1}}^{t_n}\left(1-\rho_3^{(m,j)}\right)\frac{1}{\Phi_c\left(X_s^{(j)}\right)}\Phi_c\left(X_s^{(j)}\right)\left(-\frac{1}{\sigma}\right)\Phi_{1c}\left(X_s^{(j)}\right)ds \\ &-\theta_2\sum_{n=1}^{N}\sum_{j=1}^{J}\omega_j\int_{t_{n-1}}^{t_n}\Phi_c\left(X_s^{(j)}\right)\Phi_{1c}\left(X_s^{(j)}\right)\left(-\frac{1}{\sigma}\right)X_s^{(j)}ds,\end{aligned}$$

or

$$\frac{\partial Q}{\partial c} = \frac{1}{\sigma}[- \sum_{n=1}^{N} (Z_n - Z_{n-1}) \sum_{j=1}^{J} w^{(j)} \int_{t_{n-1}}^{t_n} \left(1 - \rho_3^{(m,j)}\right) \Phi_{1c}\left(X_s^{(j)}\right) ds \\ + \theta_2 \sum_{n=1}^{N} \sum_{j=1}^{J} w^{(j)} \int_{t_{n-1}}^{t_n} \Phi_c\left(X_s^{(j)}\right) \Phi_{1c}\left(X_s^{(j)}\right) X_s{}^{(j)} ds \Bigg], \tag{4.13}$$

where $\Phi_{1c}(x) = \frac{1}{1+e^{(x-c)/\sigma}}$.

Since this latter expression cannot be solved analytically, we solve it numerically as follows: at every iteration of the MM algorithm, we use the derivative with respect to c, $\frac{\partial Q}{\partial c}$, information to move c in the right direction such that $\frac{\partial Q}{\partial c}$ goes to zero. Next, we evaluate both the model and the estimation method with some simulation examples.

4.6 Results

In this section, we explore some simulation cases to evaluate the proposed methodology to estimate parameters in a BD process with discrete time observations, and with our proposed model.

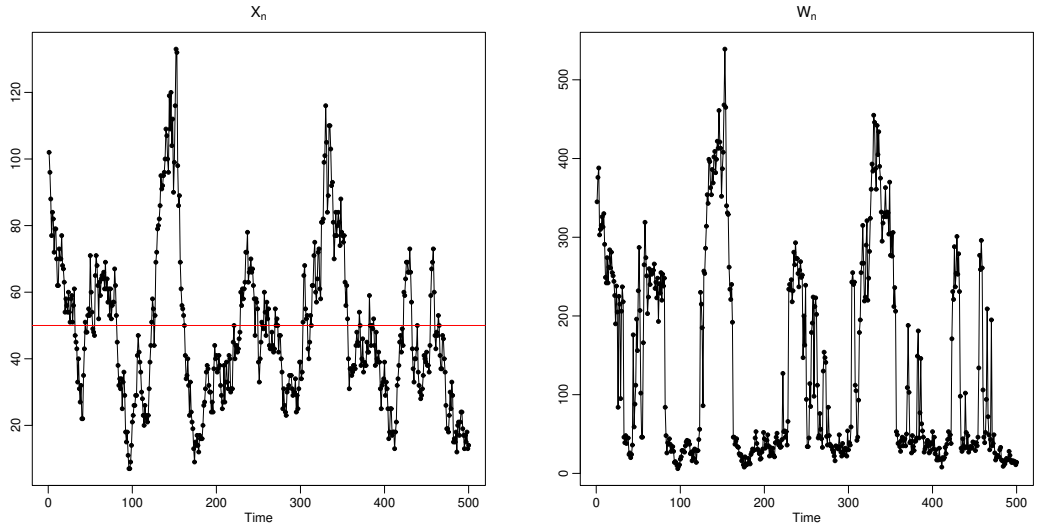
4.6.1 BD Process Example

To simulate from our model we choose birth and death rate parameters $\alpha_1 = 0.5$ and $\alpha_2 = 0.55$, and immigration and emigration rates $a = 2.2$, $b = 0$, respectively. We simulate a sample with a terminal time $N = 500$ with initial value $X_0 = 100$. The third element of the model is defined as in (4.14) with parameter values $\theta_1 = 1$, $\theta_2 = 3$, and $c = 50$. This simulation is developed by using the exact stochastic simulation algorithm of Gillespie (1977) which is discretized for $n = 1, 2, 3, ..., N$, see Wilkinson (2006) for details. Once we discretized, W_n is simulated Poisson with mean

$g(x_n)$. These discretized values represent our observation data. A realization from this simulation is displayed in Figure 4.1.

$$g(x_n) = \begin{cases} \theta_1 x_n, & \text{if } x_n \leq c \\ \theta_1 x_n + \theta_2 x_n, & \text{o.w.} \end{cases} \tag{4.14}$$

Figure 4.1: Simulation example: terminal time 500, observations period 1, X_0=100, α_1=0.5, a =2.2, α_2=0.55, $b = 0$, $\theta_1 = 1$, $\theta_2 = 3$, $c = 50$, and σ =1. The red line in figure X_n (left panel) represents the threshold parameter c.



In terms of the implementation of the proposed estimation method since we could not find a closed form for the threshold estimator, and since the implementation of any optimization procedure is highly time-consuming when the number of particles increases, we implemented the gradient ascent algorithm to get a faster estimation, see Lange (1995, 2012). This is, for the m iteration, where we use the MM algorithm, c^m is updated as $c^m = c^{m-1} + \epsilon_c \frac{\partial Q}{\partial c}$, if $\frac{\partial Q}{\partial c} < \epsilon_Q$ we keep the current value for $c^m = c^{m-1}$; ϵ_Q and ϵ_c can be defined arbitrarily, but for this example we fix them for the values $\epsilon_Q = 0.1$ and $\epsilon_c = 0.01$, respectively.

We split this simulation study into several cases to test the proposed estimation

method. a) Starting from the true parameter values, we evaluate the behavior of the estimates first for the six parameters α_1, α_2, a, θ_1, θ_2, and c; then for the five parameters by fixing the immigration rate a. We use 100 iterations with 5000 particles for the first half and 10000 particles for the second half. The tuning parameter σ is fixed at 1. To simplify the remaining simulation cases we consider the estimates for the five parameters α_1, α_2, θ_1, θ_2, and c.

b) In this case, we evaluate the convergence properties of the proposed estimators. We start some of the parameters far from the true values.

c) The effect of the tuning parameter σ is evaluated in this case. We compare the estimates by considering these values for the tuning parameter $\sigma = 0.5$, 1, 2, 4.

d) We evaluate the quality of the estimators when the time interval increase or the observations period increase. For this case, we compare the estimation behavior with observations period 0.25, 0.5, 1, 2, and 4. Also, we consider the comparison under the same number of observations and with equal terminal time.

4.6.2 Results Case a)

In Figures 4.2 and 4.3 we show the results of the estimation of each parameter, starting from the true parameter values birth rate α_1=0.5, immigration rate a =2.2, death rate α_2=0.55, emigration rate $b = 0$, infections per vector $\theta_1 = 1$, additional infection rate $\theta_2 = 3$, threshold parameter $c = 50$, initial value X_0=100 and tuning parameter σ =1. Also, the log-likelihood estimation ($loglik$), the Q function ($Qfunct$ see (4.12)), the effective sample size (EFFSS), and derivative $\frac{\partial Q}{\partial c}$ (see (4.13) are display in these plots. From these results, we conclude a fast converge for the parameters θ_1, θ_2, and c. We observe an increasing behavior regarding the estimates for the birth, immigration, and death rate parameters, although they stabilize at the end of the iterations. However, the estimation of the log-likelihood function and the Q function converge around the 20th iteration. Furthermore, the derivative $\frac{\partial Q}{\partial c}$ fluctuates around zero, which confirms the convergence of the threshold parameter c. The effective sample size (EFFSS) is close to the 50%.

In Figures 4.4 we evaluate the quality of these estimates. From the left to the right: one step ahead predictions, and residuals of X_n (top panel) and W_n (bottom panel). The one step ahead predictions are calculate with the parameter estimates of the last iteration. The terminal time is 500, with observations period 1. In Figure 4.5a black dots represent the true values for X_n and the red ones the mean of the one step ahead predictions with 10000 particles. Figure 4.5b includes the residuals. Similarly, we have in Figures 4.5c, and 4.5d for W_n. Blue dots in Figure 4.5c represents a confidence interval 3 times the standard deviation respect to the mean of the one step ahead predictions.

These outputs allow us to conclude that the proposed estimation method of particle filtering, in combination with the MM algorithm, produces good estimates for

all the parameters. Now, we will explore some of the properties of the proposed estimators using simulations.

Figure 4.2: Parameter estimates: terminal time 500, observations period 1, X_0=100, birth rate α_1=0.5, immigration rate a =2.2, death rate α_2=0.55, emigration rate $b = 0$, infections per vector $\theta_1 = 1$, additional infection rate $\theta_2 = 3$, threshold parameter $c = 50$, and tuning parameter σ =1. Log-likelihood estimation ($loglik$), Q function ($Qfunct$ see (4.12)), effective sample size (EFFSS), derivative $\frac{\partial Q}{\partial c}$ (see (4.13)).

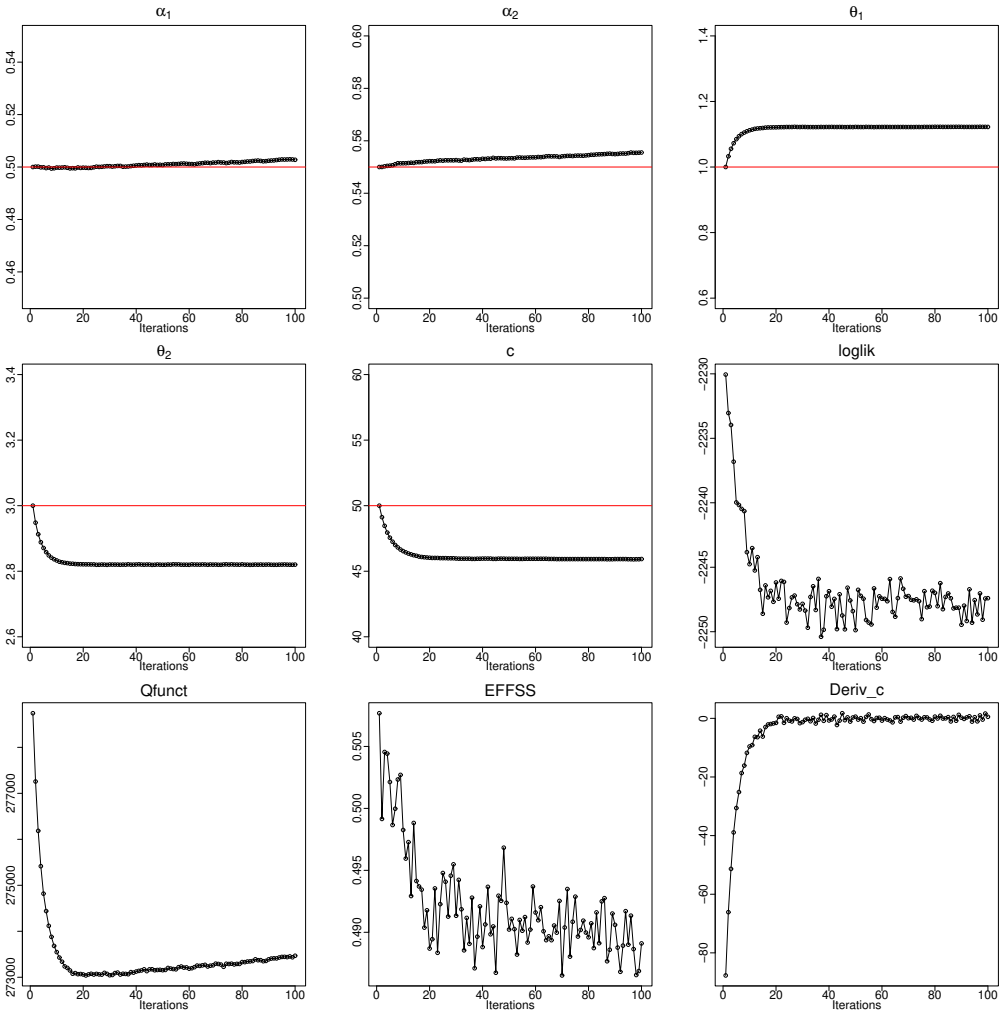


Figure 4.3: Parameter estimates: terminal time 500, observations period 1, X_0=100, birth rate α_1=0.5, immigration rate a =2.2, death rate α_2=0.55, emigration rate $b = 0$, infections per vector $\theta_1 = 1$, additional infection rate $\theta_2 = 3$, threshold parameter $c = 50$, and tuning parameter σ =1. Log-likelihood estimation ($loglik$), Q function ($Qfunct$ see (4.12)), effective sample size (EFFSS), derivative $\frac{\partial Q}{\partial c}$ (see (4.13)).

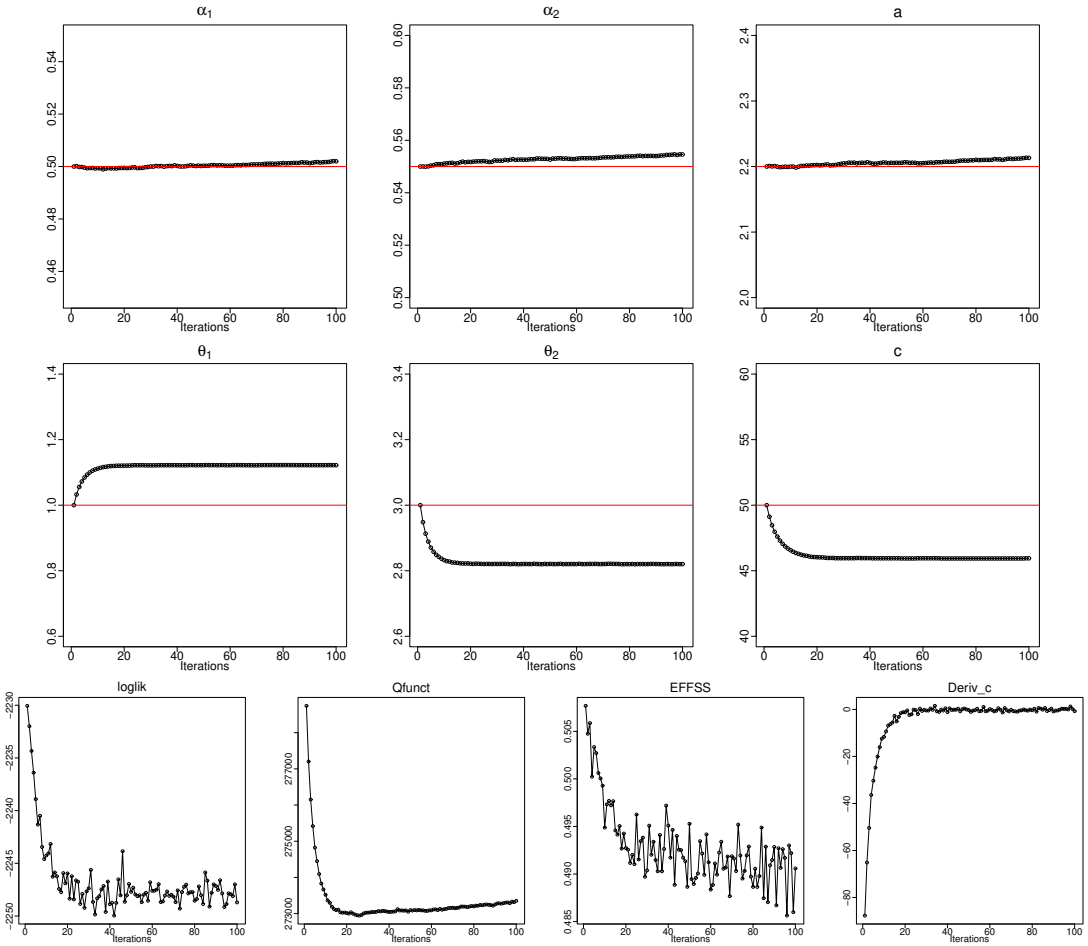
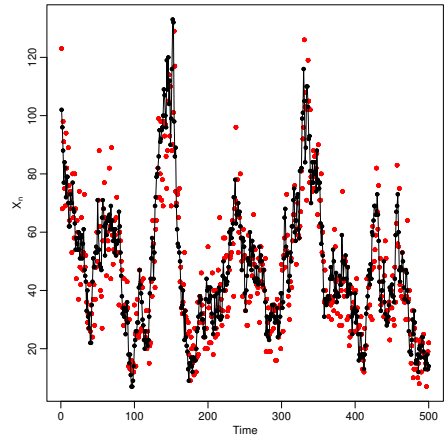
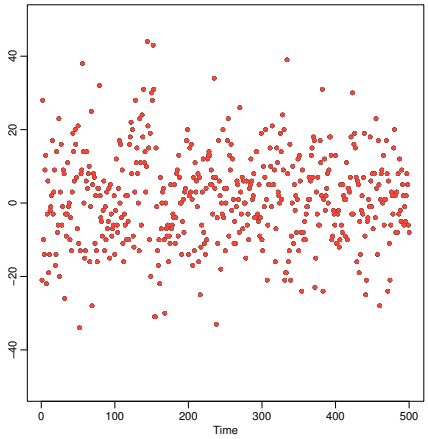


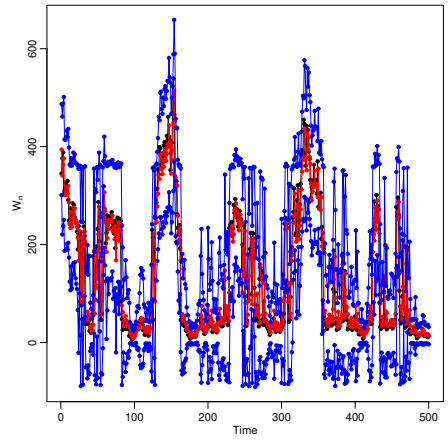
Figure 4.4: From the left to the right: one step ahead predictions, and residuals of X_n (top panel) and Wn (bottom panel) with the parameter estimates of the last iteration: terminal time 500, observations period 1, X_0=100, birth rate α_1=0.5, immigration rate a =2.2, death rate α_2=0.55, emigration rate $b = 0$, infections per vector $\theta_1 = 1$, additional infection rate $\theta_2 = 3$, threshold parameter $c = 50$, and tuning parameter σ =1.



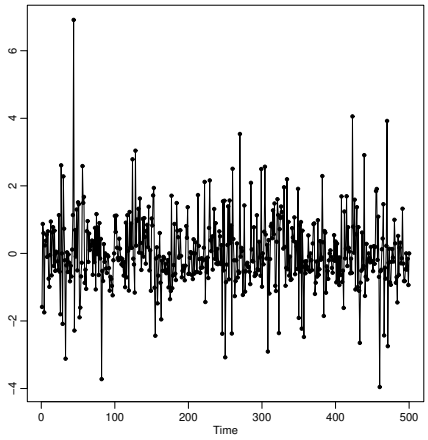
(a) One Step Ahead Predictions



(b) Residuals of X_n



(c) One Step Ahead Predictions

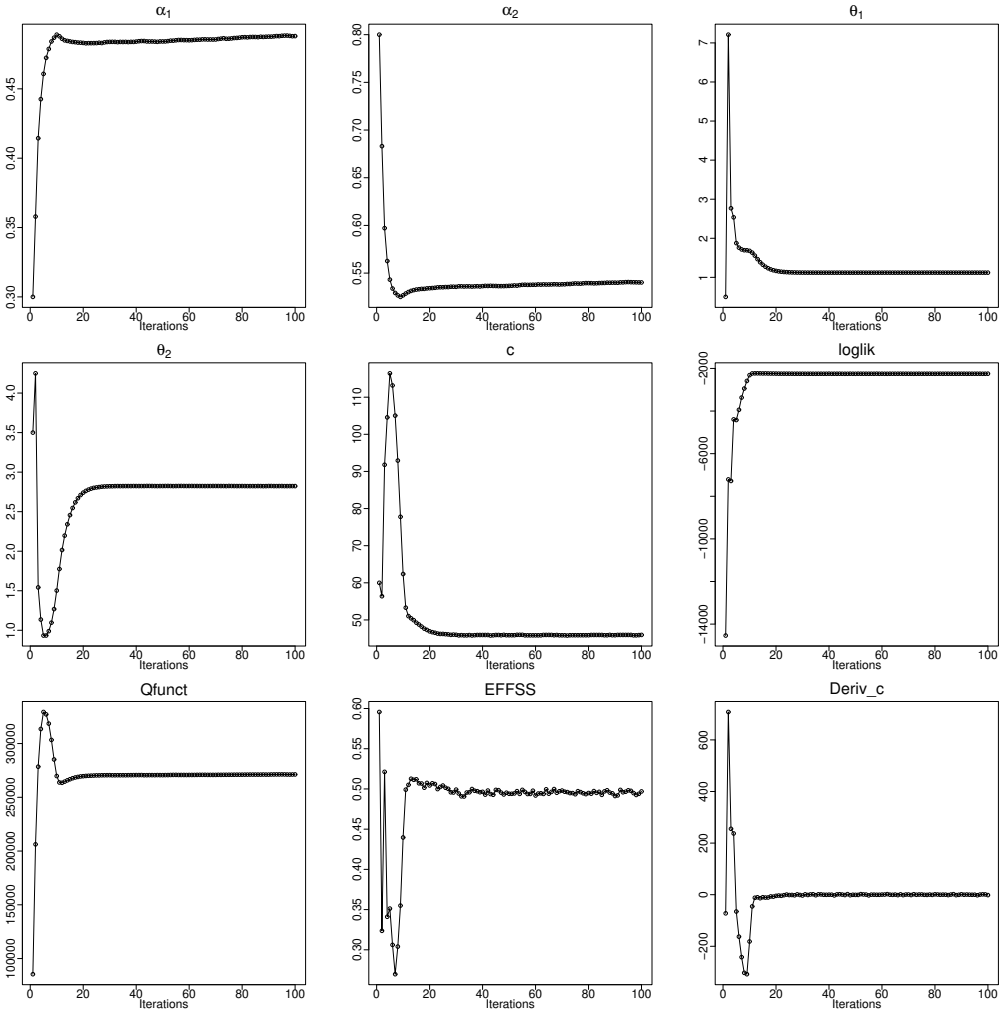


(d) Residuals of W_n

4.6.3 Results Case b)

To evaluate the robustness of convergence for these estimators, we started the estimation far away from the true values. First, we move the initial value of each parameter, one by one. Then, we move these values in pairs α_1 and α_2, θ_1 and θ_2. Finally, we move all the initial values away from the true values. Here, we only report the cases in which we move all the initial values; see Figures in 4.5. These results conclude that the convergence happens around the 20th iteration in most of the estimates.

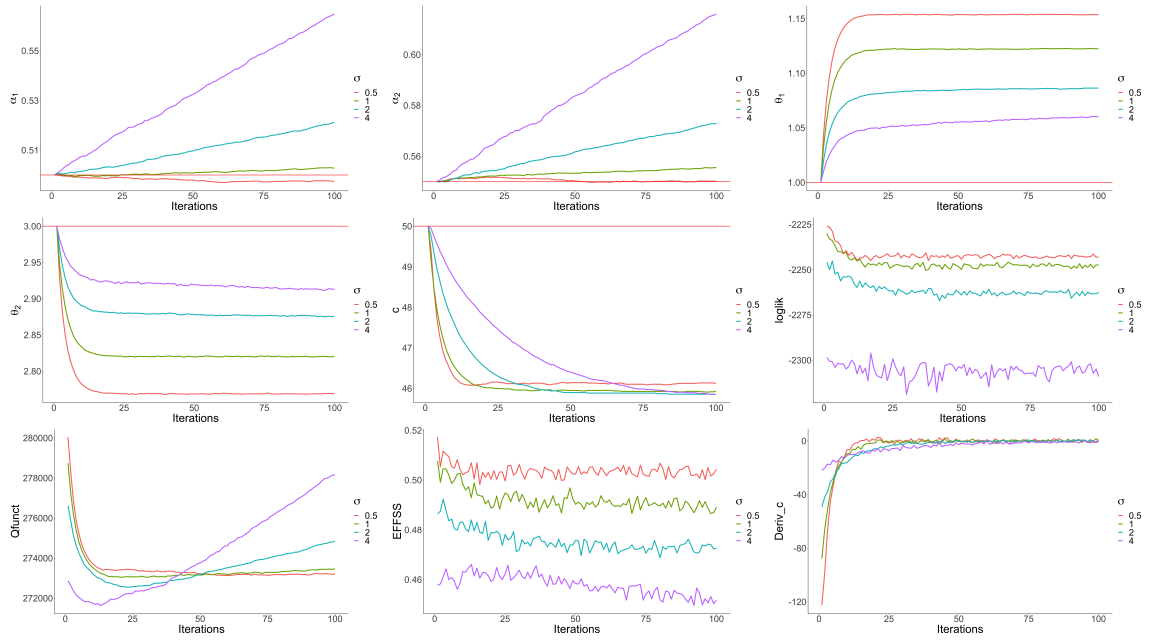
Figure 4.5: Parameter estimates starting far away from the true values: terminal time 500, observations period 1, X_0=100, birth rate α_1=0.5, immigration rate a =2.2, death rate α_2=0.55, emigration rate $b = 0$, infections per vector $\theta_1 = 1$, additional infection rate $\theta_2 = 3$, threshold parameter $c = 50$, and tuning parameter σ =1. Log-likelihood estimation ($loglik$), Q function ($Qfunct$ see (4.12)), effective sample size (EFFSS), derivative $\frac{\partial Q}{\partial c}$ (see (4.13)).



4.6.4 Results Case c)

This case evaluates the effect of the tuning parameter in the estimation procedure. Figure 4.6 shows the results of the estimation as well as the checking plots. In these outcomes, the convergence for the estimates of the parameters θ_1. θ_2 and c is convincing. For parameters θ_1. θ_2 as the tuning parameter increases, the estimates move close to the true value. Regarding the threshold parameter c, the estimates converge for all the cases, but the convergence demands more iterations as the tuning parameter increase. We observe similar behavior for the estimates of the birth and death rate parameters α_1 and α_2, i.e., as the value of the tuning parameter increase, we need to increase the number of iterations. Still, the estimates go far away from the actual values. The log-likelihood estimate and the Q function, the estimates with high values of σ are not stable. Also, the effective sample size tends to decrease as the tuning parameter increases. In conclusion, these results suggest small values for σ close to 1 or 2 at least given the other parameters values.

Figure 4.6: Parameter estimates using different values for the tuning parameter $\sigma =$ 0.5, 1, 2, 4, 8.: terminal time 500, observations period 1, X_0=100, birth rate α_1=0.5, immigration rate a =2.2, death rate α_2=0.55, emigration rate $b = 0$, infections per vector $\theta_1 = 1$, additional infection rate $\theta_2 = 3$, and threshold parameter $c = 50$. Log-likelihood estimation ($loglik$), Q function ($Qfunct$ see (4.12)), effective sample size (EFFSS), derivative $\frac{\partial Q}{\partial c}$ (see (4.13)).



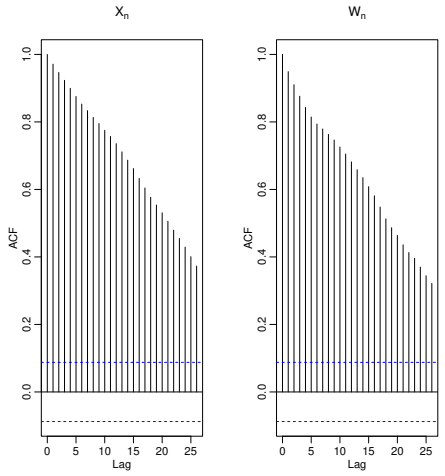
4.6.5 Results Case d)

This case evaluates the quality of the proposed estimators in terms of the observation period and the terminal time. Firstly, this case evaluates the effect of scale separation between the vector dynamics and the case observations as the time between observations grows. Let $\Delta = t_{nOP} - t_{(n-1)OP}$ the separation between the vector dynamics, where OP is the observation period, and $n \in \mathbb{N}$; therefore, we have observations every OP time unit. Thus, as OP increases, both the X_n (we keep the notation X_n to notice the discretization of the true latent process) and W_n tend to be uncorrelated. To confirm this fact, we simulate X_t by following the proposed model with these conditions: initial value X_0=100, birth rate α_1=0.5, immigration rate a =2.2, death rate α_2=0.55, emigration rate $b = 0$, infections per vector $\theta_1 = 1$, additional infection rate $\theta_2 = 3$, and threshold parameter $c = 50$. The values we selected for the observation period are $\{0.5, 1, 2, 4\}$. Then we create the ACF plots for X_n, which is the discretization of the true latent process, and W_n the observations. Figures 4.7 include the ACF plots for X_n and W_n for the selected observation period (OP) values, and by increasing the terminal time (TT), in order to have the same number of observations in each case, $TT = 500$. From these outcomes, we confirm that as the OP increases, the autocorrelations of X_n and W_n decrease. Also, in Figures 4.8 we can see how the estimates move far away from the true values, and the effective sample size also decreases as the OP increases.

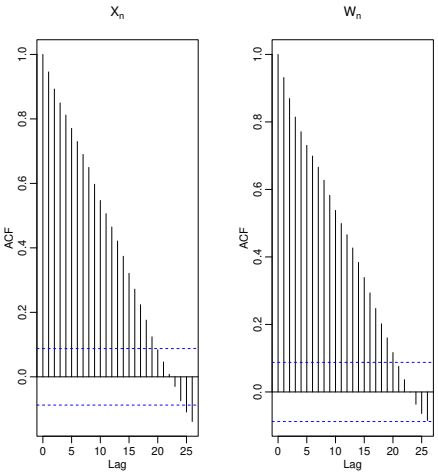
Furthermore, we evaluate the estimates' behavior when TT increases by keeping the same OP selected values. In this example, the ACF plots have similar behavior as the previous ones, see Figures 4.9. The results in Figures 4.10 show that as the OP increases the estimates for the parameters θ_1, θ_2, and c move far from the true values; however, the estimates for α_1 and α_2 do not exhibit a convergence behavior. In

addition, we isolate the effect of the threshold parameter by fixing it in the estimation process. By fixing c, the convergence of the parameters α_1 and α_2 improve for $OP = 1$, see Figures 4.11. From these results, we conclude that as the number of observations increases, more iterations are needed, at least for the two parameters α_1 and α_2; however, it becomes highly computational and time-consuming.

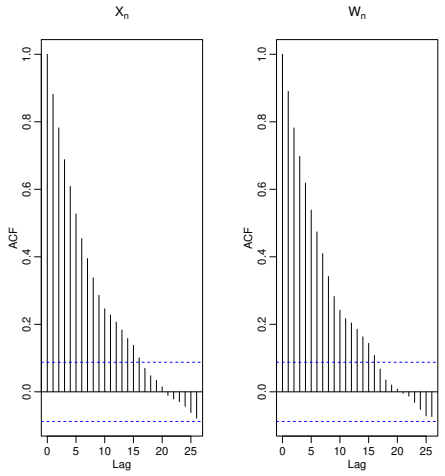
Figure 4.7: ACF plots (using different observations period (OP) 0.5, 1, 2, 4 and different terminal time (TT) 250, 500, 1000, 2000, respectively; for a total of 500 observations for each scenery. Initial value X_0=100, birth rate α_1=0.5, immigration rate a =2.2, death rate α_2=0.55, emigration rate $b = 0$, infections per vector $\theta_1 = 1$, additional infection rate $\theta_2 = 3$, and threshold parameter $c = 50$.



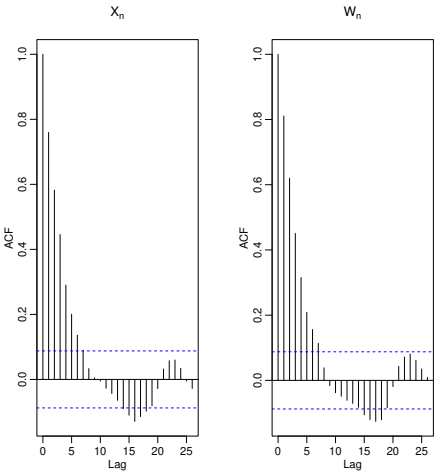
(a) $OP = 0.5$, $TT = 250$



(b) $OP = 1$, $TT = 500$



(c) $OP = 2$, $TT = 1000$



(d) $OP = 4$, $TT = 2000$

Figure 4.8: Parameter estimates using different observations period (OP) 0.5, 1, 2, 4 and different terminal time (TT) 125, 250, 500, 1000, 2000, respectively; for a total of 500 observations for each scenery. Initial value X_0=100, birth rate α_1=0.5, immigration rate a =2.2, death rate α_2=0.55, emigration rate b = 0, infections per vector θ_1 = 1, additional infection rate θ_2 = 3, and threshold parameter c = 50. Log-likelihood estimation ($loglik$), Q function ($Qfunct$ see (4.12)), effective sample size (EFFSS), derivative $\frac{\partial Q}{\partial c}$ (see (4.13)).

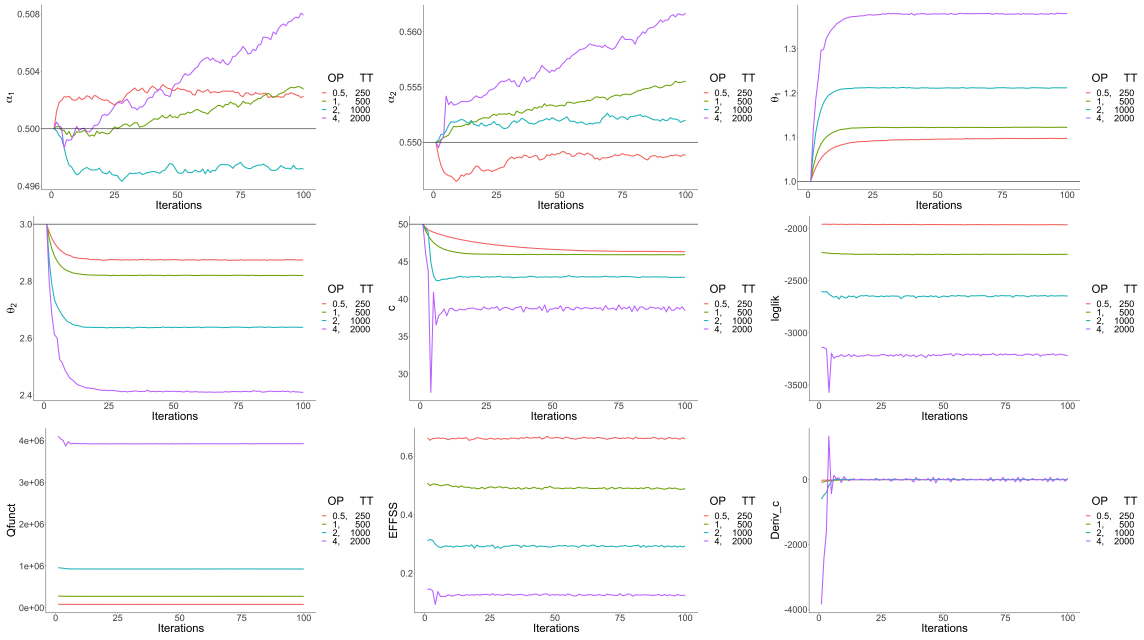
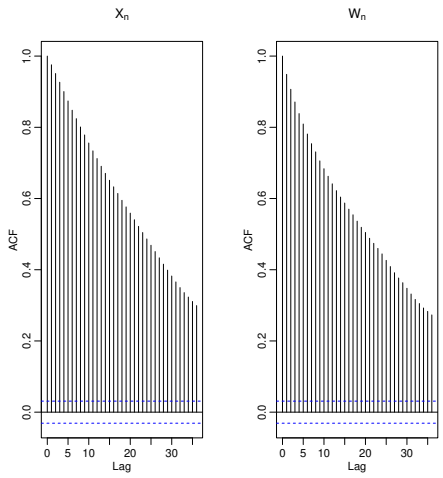
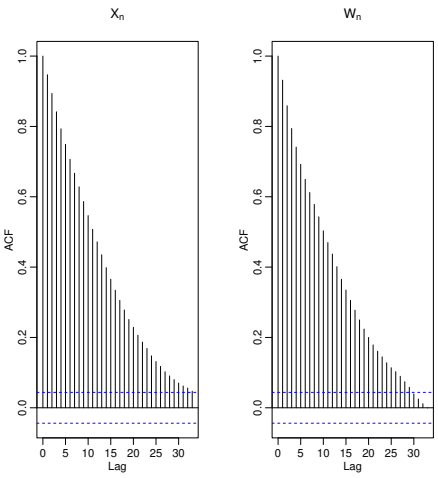


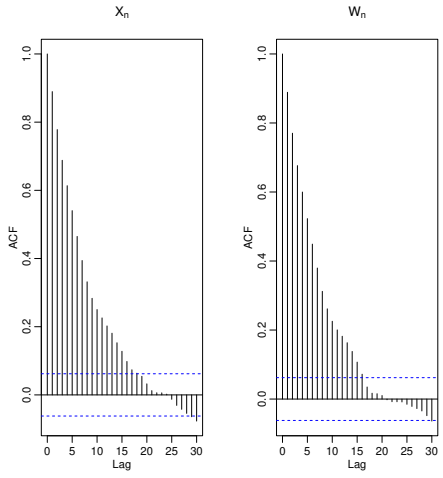
Figure 4.9: ACF plots using different observations period (OP) 0.5, 1, 2, 4 and a fix terminal time (TT) 2000. Initial value X_0=100, birth rate α_1=0.5, immigration rate a =2.2, death rate α_2=0.55, emigration rate $b = 0$, infections per vector $\theta_1 = 1$, additional infection rate $\theta_2 = 3$, and threshold parameter $c = 50$.



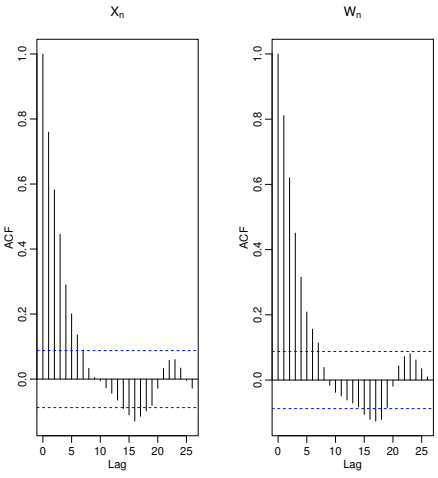
(a) $OP = 0.5$



(b) $OP = 1$



(c) $OP = 2$



(d) $OP = 4$

Figure 4.10: Parameter estimates using different observations period (OP) 0.5, 1, 2, 4 and fix terminal time 2000. Initial value X_0=100, birth rate α_1=0.5, immigration rate a =2.2, death rate α_2=0.55, emigration rate $b = 0$, infections per vector $\theta_1 = 1$, additional infection rate $\theta_2 = 3$, and threshold parameter $c = 50$. Log-likelihood estimation ($loglik$), Q function ($Qfunct$ see (4.12)), effective sample size (EFFSS), derivative $\frac{\partial Q}{\partial c}$ (see (4.13)).

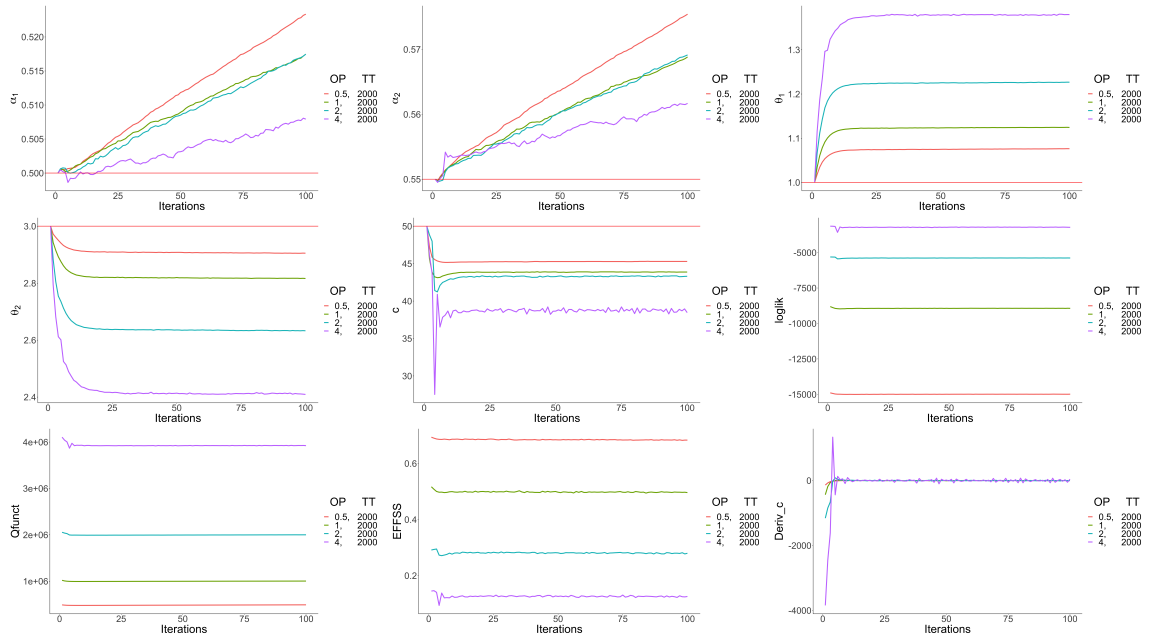
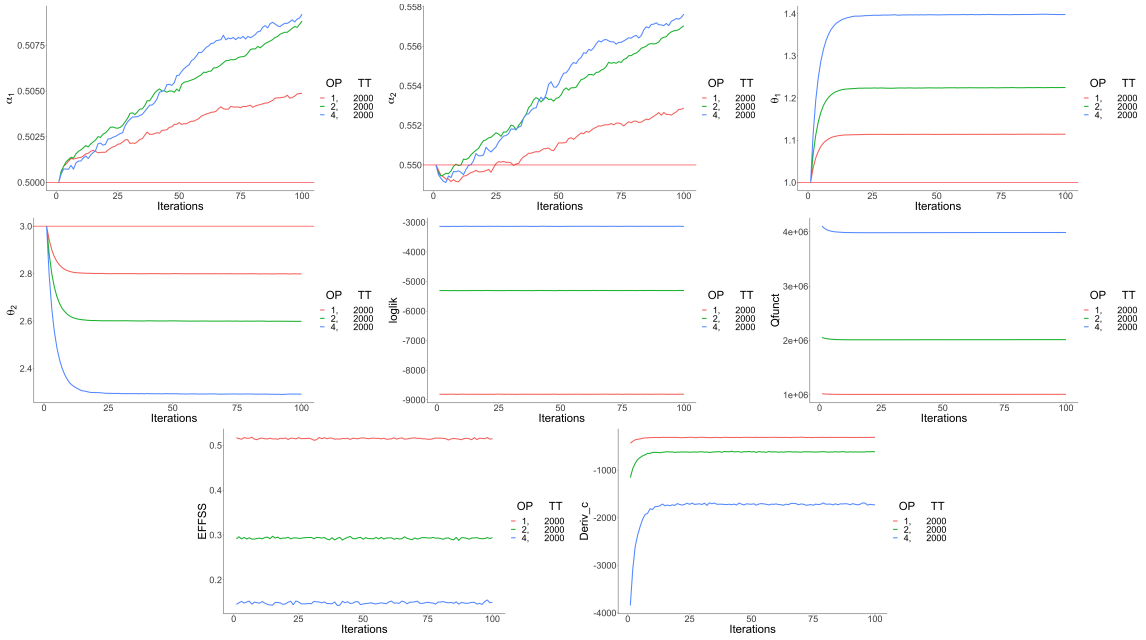


Figure 4.11: Parameter estimates using different observations period (OP) 1, 2, 4 and fix terminal time (TT) 2000. Initial value X_0=100, birth rate α_1=0.5, immigration rate a =2.2, death rate α_2=0.55, emigration rate $b = 0$, infections per vector $\theta_1 = 1$, additional infection rate $\theta_2 = 3$, and threshold parameter is fixed at $c = 50$. Log-likelihood estimation ($loglik$), Q function ($Qfunct$ see (4.12)), effective sample size (EFFSS), derivative $\frac{\partial Q}{\partial c}$ (see (4.13)).



4.7 Case Study

Cholera is a disease caused by infection of the small intestine by Vibrio cholera characterized by massive diarrhea, vomiting, and dehydration: death occurs in severe untreated cases. Mainly, it is endemic in easter India, and Bangladesh (Rabbani and Greenough (1999)). The world experienced seven cholera pandemics in the last 194 years. Outbreaks can occur where water supply, sanitation, food safety, and hygiene are inadequate. Also, contaminated food (especially seafood) is a more common cause of cholera in developed countries, whereas contaminated water is more common in developing countries.

Early observations have recognized that fish and shellfish are important vehicles of the transmission of cholera. This kind of seafood has been the culprit in many cholera outbreaks since the nineteenth century. When the surrounding water is contaminated by sewage or other environmental sources, fish are likely contaminated by cholera. In Dhaka, the capital city o Bangladesh two outbreaks of cholera in 1974 and 1975 were identified Khan *et al.* (1983). A case-control study revealed that the attacks of cholera were significantly associated with eating seafood in restaurants.

The study of the dynamics of cholera in Dhaka has been in the literature in the last several decades. Koelle and Pascual (2004) proposes a non-linear time series model which include extrinsic effects such as seasonality and long-term changes, for example. In this section, we will extend the model proposed in section 4.4 in the context of cholera deaths reported in Bengal's Dacca district from 1891 to 1940, see Ionides *et al.* (2015). This data set presents a seasonal behavior that is considered in the specification of the birth-death model by means of a sinusoidal function. Here we also adapted the estimators, particularly the estimator for the birth parameter α_1.

4.7.1 Model Extension

The extension of the model in (4.1) consists in a simple modification in the birth term, this is

$$X_t = X_0 + Y_1\bigg(\int_0^t (\alpha_{1s}X_s + a)ds\bigg) - Y_2\bigg(\int_0^t (\alpha_2 X_s + b)I_{\{X_s>0\}}ds\bigg), \tag{4.15}$$

where

$$\alpha_{1s} = \beta_1\bigg(\cos\bigg(\frac{4\pi}{12}s + \omega\bigg) + 1\bigg), \tag{4.16}$$

$$\begin{aligned} g(x_t) &= \theta_1 x_t + \theta_2 x_t I_{\{X_t>c\}}, \\ Z_t &= Y_3\bigg(\int_0^t g\big(X_s\big)ds\bigg), \end{aligned} \tag{4.17}$$

where $I_{\{X_t>c\}} = 1$ if $X_t > c$ and zero otherwise; and

$$W_n = Z_{t_n} - Z_{t_{n-1}}. \tag{4.18}$$

4.7.2 Inference

By following the estimation procedure exposed in section 4.5.1, the log-likelihood function is given in terms of the parameters birth rate α_{1s} through the parameter β_1 (we assume that ω is known), immigration rate a, death rate α_2, emigration rate b, and the parameters of the observation equation θ_1, θ_2, and the threshold c. The estimators for all the parameters except for β_1 and a are the same as we found in section 4.5.1. Then we will calculate the estimators for these.

Replacing the birth rate term in the log-likelihood function in (4.6) and by applying the minorize-maximize (MM) algorithm, Lange (2012) over the first term we have:

$$\rho = \rho_{1*}^{(m)} = \frac{\alpha_{1s}^{(m)} X_s}{\alpha_1^{(m)} X_s + a^{(m)}}. \tag{4.19}$$

Replacing α_{1s} by the expression in (4.16) we obtain

$$\rho = \rho_{1*}^{(m)} = \frac{\beta_1^{(m)}\big(\cos\big(\frac{4\pi}{12}s + \omega\big) + 1\big)X_s}{\beta_1^{(m)}\big(\cos\big(\frac{4\pi}{12}s + \omega\big) + 1\big)X_s + a^{(m)}}. \tag{4.20}$$

Thus, the estimators for β_1 and a are given by

$$\hat{\beta}_1^{(m+1)} = \frac{\sum_{n=1}^{N}\sum_{j=1}^{J} w^{(j)} \int_{t_{n-1}}^{t_n} \rho_{1*}^{(m,j)} dU_1^{(j)}(s)}{\sum_{n=1}^{N}\sum_{j=1}^{J} \omega^{(j)} \int_{t_{n-1}}^{t_n} \big(\cos\big(\frac{4\pi}{12}s + \omega\big) + 1\big)X_s^{(j)} ds},$$

and

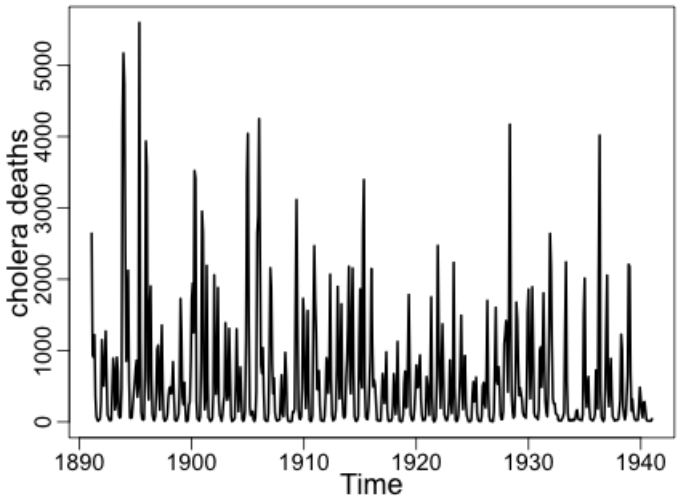
$$\hat{a}^{(m+1)} = \frac{\sum_{n=1}^{N}\sum_{j=1}^{J} w^{(j)} \int_{t_{n-1}}^{t_n} (1 - \rho_{1*}^{(m,j)}) dU_1^{(j)}(s)}{N},$$

respectively. Now we will fit the extended model to the cholera deaths reported in Bengal's Dacca district and the corresponding parameter estimates.

4.7.3 Results and Discussions

Ionides *et al.* (2015) analyze the observed Cholera mortality data from Bengal's Dacca district over the years 1891 to 1940, see Figure 4.12. The data, consisting of monthly counts of cholera mortality, is available in the R package *pomp* (King *et al.* (2016)). In Ionides *et al.* (2015), based on the model developed in King *et al.* (2007), an extension of the Susceptible-Infected-Recovered (SIR) model is proposed. This model, includes the number of births, related to the total population size in the susceptible component of the SIR model. Also, this model includes seasonal transmission rates and seasonal environmental reservoir parameters; each of these seasonal effects includes 6 components or parameters.

Figure 4.12: Cholera deaths reported in Bengal's Dacca district over the years 1891 to 1940



(a)

To fit the proposed model to the Cholera mortality data, we do an exploration of the model by defining a grid of values for the parameters β_1, α_2, θ_1, θ_2 and the threshold parameter c. Based on this exploration we established the following initial conditions: initial value $X_0 = W_1/50$, where W_1 is the first observation Cholera mortality, birth rate β_1=0.5, immigration rate a =0, death rate α_2=0.501, emigration rate $b = 0$, infections per vector $\theta_1 = 1$, additional infection rate $\theta_2 = 50$, threshold parameter $c = 20$, and $\omega = -2$. Thus, we run 100 iterations with the MM algorithm with 5000 particles in the first half of the iterations and 1000 particles in the second half.

Figures 4.13 show the results of the parameter estimates. All the estimates exhibit a convergence behavior around 0.7 for the parameters β_1 and α_2, approximately 11 for the parameter θ_1, close to 0.5 for the parameter θ_2 and around 80 for the threshold parameter c. Both the Log-likelihood estimation ($loglik$) and Q function ($Qfunct$)

stabilize around the 15 iterations, and similarly happen with the derivative $\frac{\partial Q}{\partial c}$ whose values fluctuate around zero after the ten first iterations. The effective sample size (EFFSS) moves around 20%. From these results, although we observed a convergence behavior the estimates for the parameter θ_2 is effectively zero and the effective sample size is essentially zero.

About the quality of these estimates we include in Figures 4.14 the one step ahead prediction and the corresponding residuals. Although the predictions try to track the data's dynamic in general the predictors are poor.

4.7.4 Other Extensions

Also, using the diffusion representation of the birth-death process (Adke and Moyal (1963); Pinto *et al.* (2009)), we could use the proposed model to analyze the number of investors in the financial system. Further it would be necessary to modify the distribution of the observation equation where the data would include the asset price at the end of each trading day. A third potential extension is the Covid cases where the observations are Covid deaths. In this context, when the number of cases goes above a certain threshold, we could expect more deaths which implies a change in the death rate. That is because the hospitals are full, and the capacity to care for people exceeds their limits. This situation was observed in several stages of the pandemic around the US and the world. SIS epidemic model would be a good representation of this extension, see Crawford *et al.* (2014).

Figure 4.13: Parameter estimates initial value X_0=52.82, birth rate β_1=0.5, immigration rate a =0, death rate α_2=0.501, emigration rate $b = 0$, infections per vector $\theta_1 = 1$, additional infection rate $\theta_2 = 50$, threshold parameter $c = 20$. Log-likelihood estimation ($loglik$), Q function ($Qfunct$ see (4.12)), effective sample size (EFFSS), derivative $\frac{\partial Q}{\partial c}$ (see (4.13)).

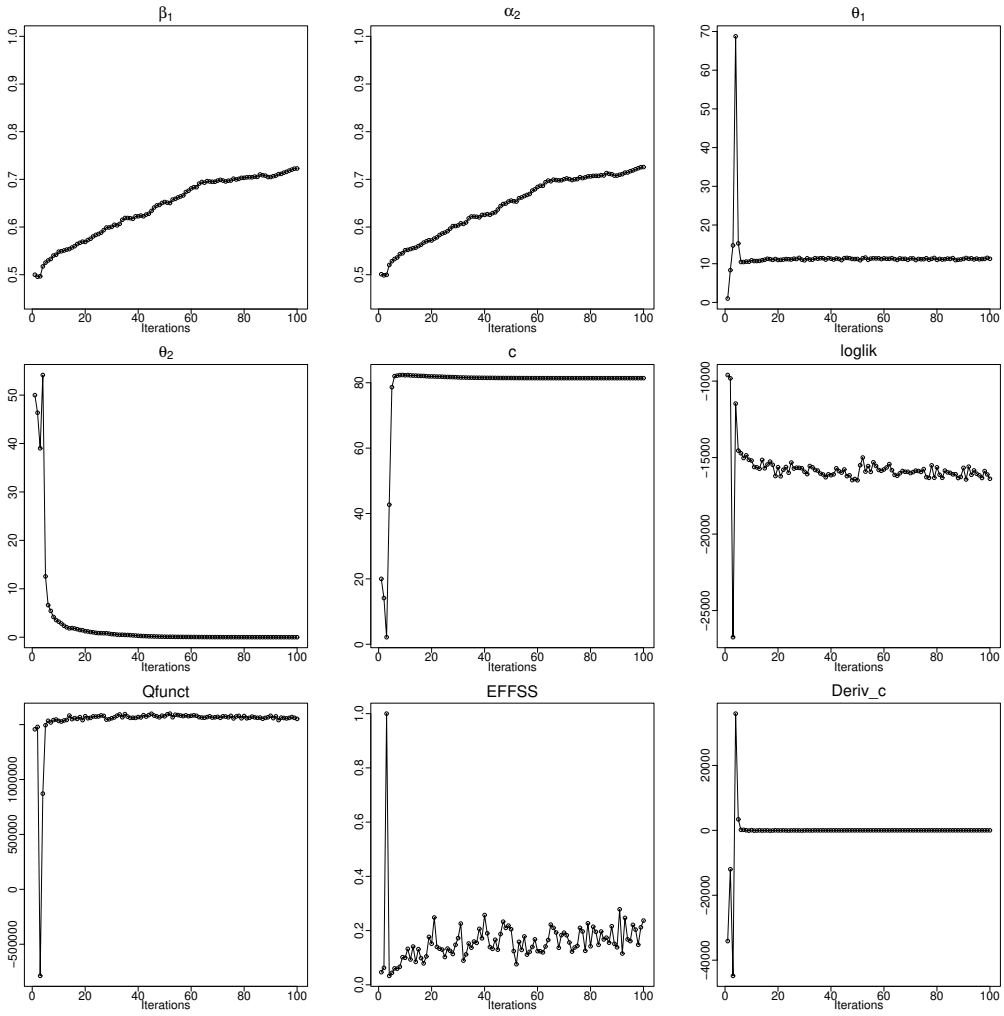
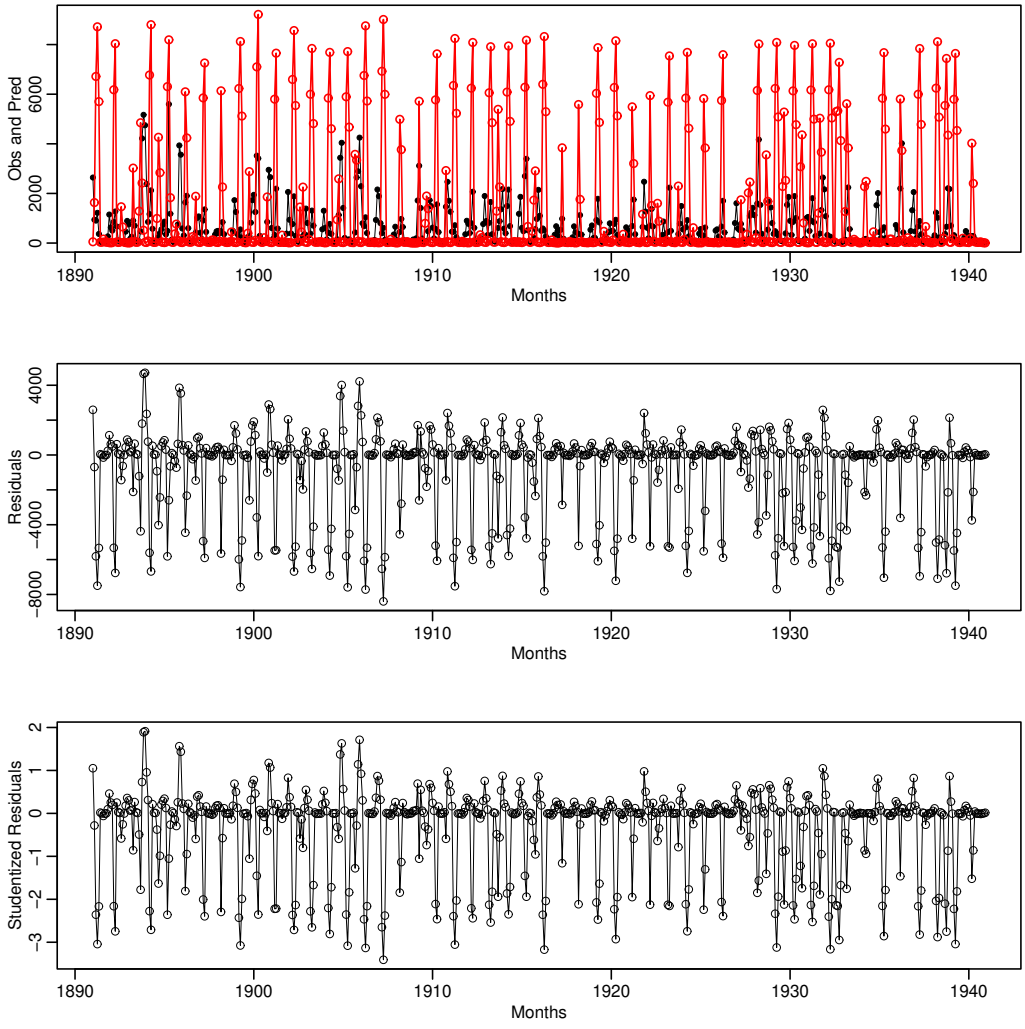


Figure 4.14: Top panel: Cholera mortality data (black dots) and one step ahead predictions (red dots). Middle panel: Residuals of Cholera mortality data. Bottom panel: Studentized Residuals of Cholera mortality data



4.8 Conclusion

This study proposes a new model for disease incidence that evaluates the interaction between two species which incorporates a birth-death process to model the unobserved population size of a vector species. In addition, the transmission rate from the vector species to the second one includes a threshold parameter that measures the effect of the transmission when the host species' population size surpasses a certain level. Our proposed model was evaluated under some simulation cases. Also, we offer a novel estimation method that combines particle filter with the MM algorithm. The particle filter allows us to recreate paths in the continuous-time domain, representing the host population dynamic. Thus, we calculated weights for each path and each time interval given by the second species' incidence reports collected in discrete time. As a consequence, these calculations were used in the defined estimators. The proposed estimators are an extension of the ones proposed by Crawford *et al.* (2014).

The simulation studies reveal that our model and our estimation method behave well. Four simulation cases were considered. The first case allowed us to conclude that the proposed estimation method of particle filtering, in combination with the MM algorithm, produces good estimates for all the parameters, even when we included the immigration rate parameter. Also, the quality of our estimates was assessed by contrasting the simulated observations with the one step ahead predictions. We confirmed that the estimates behaved well according to the residuals' behavior. The second case, allowed us to conclude about the robustness of the convergence. In general, the convergence happens around the 20th iteration in most of the estimates.

Since the proposed observation model includes an indicator function, we created an approximation of the model using a logistic function to ensure that the likelihood

function is continuous and differentiable in the threshold parameter c. The defined logistic function has a tuning parameter σ, therefore, the third simulation case assessed the effect of this tuning parameter. At least for the parameter values considered in this simulation case, small values for σ close to 1 or 2 are the best choice. The fourth case evaluated the effect of the observation period (OP), concluding that as the OP increases, the estimates for all the parameters move far from the true value. However, more iterations are necessary for the estimation process when the number of observations increases.

We adapted the proposed model to the Cholera mortality data from Bengal's Dacca district. Although this model is simple compared with the model fitted in Ionides *et al.* (2015), this tries to follow the movements of the data. The parameter estimates show good behavior regarding their convergence; however, the estimates for the parameter θ_2 are close to zero. According to the low effective sample size, the one step ahead predictors are poor.

Although the proposed model did not fit properly with the Cholera mortality data, some conditions could be considered to improve the fitted model. For example, by defining some restrictions over the parameter θ_2. Furthermore, a simulation study over this case could be considered, especially when searching the initial condition for the hidden states and initial values for the parameters in terms of the scaling. Also, including more seasonal components, or even covariates, could be an interesting improvement of the model. Finally, we recognize that the dynamic of these data is complex, and maybe this data set is not the best choice to assess the proposed model. Other examples could be considered in future works like the extensions we mentioned above, such as in the context of COVID cases or financial data.

Chapter 5

CONCLUSION

This dissertation explores a previously proposed non-linear state-space model to estimate measles transmission at the country level. A Bayesian approach that uses particle MCMC (pMCMC) was developed to estimate the parameters of the non-linear state-space model. Also, iteration filtering (IF2) was used as a support method to verify the Bayesian estimation and inform prior distributions' selection. Particle Markov Chain Monte Carlo (pMCMC) and Iteration Filtering (IF) are complex computational methods based on the Bayesian framework and likelihood maximization, respectively. Thus, Chapter 2 provided a background of the pMCMC and the iteration filtering (IF2). Chapter 3 implemented these computational tools that provided a novel set of estimators for the non-linear state-space model.

In particular, Chapter 2 presented the Bayesian estimation that uses pMCMC and iteration filtering by using two approaches: one by following the R package *pomp* and a second by using our implementation. The two approaches were contrasted via simulation, obtaining similar results and running time for the performance of the particle filtering procedure; however, the time expended in the complete adaptation for both pMCMC and IF2 was superior in our implementation.

Chapter 3 presented a non-linear state-space model to estimate measles transmission at the country level as proposed in Eilertson et al. (2019) and provided a presentation of a method to perform Bayesian analysis of the whole country level. The exploration of the model and the implementation of the Bayesian methods were carried out firstly by developing some simulation studies and secondly by considering some case studies. Addressing simulation examples with three and then four pa-

rameters for the model allowed the evaluation of the proposed Bayesian estimation procedure. The case of the fourth parameter allowed us to find that the estimates of the posterior density of the σ^2 parameter is quite sensitive to the choice of priors. Regarding this fact, these results mimic the ones found in previous studies. Also, this chapter evaluated the model by performing one-step ahead and forward predictions. The calculation of these predictions was based on three methods and the combination of different selected priors and proposals, allowing a variety of predictions and mainly confirming the robustness of the implemented Bayesian analysis. The case studies exploited the benefit of the faster computational properties of the IF2 algorithm to find reasonable initial conditions S0, initial parameter values, and priors to perform the estimations using the pMCMC algorithm. The result from this application allowed confirmation of the robustness of the method and also allowed identification of some weaknesses of the method in the context of countries with few reported cases and countries with extreme observations. Thus, the provided method improves on the work in Eilertson et al. (2019) with a lower computational cost.

Finally, Chapter 4 proposed a birth-death process to model the unobserved population size of a disease vector. This model studies the effect of a disease vector population size on a second affected population. There are many similar models, but the presentation given here of a case number that depends on a birth-death process is unique. Also, this chapter reformulated and extended the model presented in Crawford et al. (2014), making the proposed likelihood method much more compatible with particle filters. The proposed model was evaluated under some simulation cases. Also, a novel estimation method was offered that combines particle filter with the MM algorithm. In addition, the first steps of a case study are in progress, extending the proposed model.

Adke, S. and J. Moyal, "A birth, death, and diffusion process", Journal of Mathematical Analysis and Applications **7**, 2, 209–224 (1963).

Akman, V. E. and A. E. Raftery, "Asymptotic Inference for a Change-Point Poisson Process", The Annals of Statistics **14**, 4, 1583 – 1590 (1986).

Allen, L. J. S., *An Introduction to Stochastic Epidemic Models*, pp. 81–130 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2008).

Anderson, R. M. and R. M. May, *Infectious Diseases of Humans: Dynamics and Control*. (Oxford: Oxford University Press, 1991).

Andrieu, C., A. Doucet and R. Holenstein, "Particle markov chain monte carlo methods", Journal of the Royal Statistical Society Series B **72**, 3, 269–342 (2010).

Aue, A. and L. Horváth, "Structural breaks in time series", Journal of Time Series Analysis **34**, 1–16 (2013).

Behrens, C. N., H. F. Lopes and D. Gamerman, "Bayesian analysis of extreme events with threshold estimation", Statistical Modelling **4**, 3, 227–244 (2004).

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

Bollerslev, T., "Generalized autoregressive conditional heteroskedasticity", Journal of Econometrics **31**, 3, 307–327 (1986).

Boudjellaba, H., B. MacGibbon and P. Sawyer, "On exact inference for change in a poisson sequence", Communications in Statistics - Theory and Methods **30**, 3, 407–434 (2001).

Chen, S., J. Fricks and M. J. Ferrari, "Tracking measles infection through non-linear state space models", Journal of the Royal Statistical Society. Series C (Applied Statistics) **61**, 1, 117–134 (2012).

Crawford, F. W., V. N. Minin and M. A. Suchard, "Estimation for general birth-death processes", Journal of the American Statistical Association **109**, 506, 730–747 (2014).

Crawford, F. W. and M. A. Suchard, "Transition probabilities for general birth-death processes with applications in ecology, genetics, and evolution.", J Math Biol **65**, 3, 553–580 (2012).

Dahlin, J. and T. B. Schön, "Getting started with particle metropolis-hastings for inference in nonlinear dynamical models", Journal of Statistical Software **88** (2019).

Davison, A. C. and R. L. Smith, "Models for exceedances over high thresholds", Journal of the Royal Statistical Society. Series B (Methodological) **52**, 3, 393–442 (1990).

de Valpine, P., D. Turek, C. Paciorek, C. Anderson-Bergman, D. Temple Lang and R. Bodik, "Programming with models: writing statistical algorithms for general model structures with NIMBLE", Journal of Computational and Graphical Statistics **26**, 403–417 (2017).

Dempster, A. P., N. M. Laird and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm", Journal of the Royal Statistical Society, Series B **39**, 1, 1–38 (1977).

Diebolt, J., M. El-Aroui, M. Garrido and S. Girard, "Quasi-conjugate bayes estimates for gpd parameters and application to heavy tails modelling", Extremes **1**, 57–78 (2005).

do Nascimento, F., D. Gamerman and H. Lopes, "A semiparametric bayesian approach to extreme value estimation", Statistics and Computing **22**, 661–675 (2012).

Douc, R., E. Moulines and D. Stoffer, *Nonlinear Time Series: Theory, Methods and Applications with R examples.* (Taylor & Francis, CRC Press, Boca Raton, Fl., 2014).

Douc, R., E. Moulines and D. Stoffer, *Non Linear Time Series (Theory, Methods, and Applications with R Examples.* (A Chapman & Hall Book. Test in Statistical Science, 2017).

Doucet, A., N. de Feitas and N. Gordon, *Sequential Monte Carlo Methods in Practice* (Springer-Verlag, 2010).

Drovandi, C. and A. Pettitt, "Estimation of parameters for macroparasite population evolution using approximate bayesian computation", Biometrics **67**, 1, 225—233 (2011).

Eilertson, K. E., J. Fricks and M. J. Ferrari, "Estimation and prediction for a mechanistic model of measles transmission using particle filtering and maximum likelihood estimation.", Statistics in Medicine **38**, 4146–4158 (2019).

Endo, A., E. van Leeuwen and M. Baguelin, "Introduction to particle markov-chain monte carlo for disease dynamics modellers", Epidemics **29**, 100363 (2019).

Frigessi, A., O. Haug and H. Rue, "A dynamic mixture model for unsupervised tail estimation without threshold selection", Extremes **5**, 219–235 (2002).

Gillespie, D. T., "Exact stochastic simulation of coupled chemical reactions", The Journal of Physical Chemistry **81**, 2340–2361 (1977).

Ho, L. S. T., J. Xu, F. W. Crawford, V. N. Minin and M. A. Suchard, "Birth/birth-death processes and their computable transition probabilities with biological applications", Journal of Mathematical Biology **76**, 911–944 (2018).

Ionides, E. L., C. Bretó and A. A. King, "Inference for nonlinear dynamical systems", Proceedings of the National Academy of Sciences **103**, 49, 18438–18443 (2006).

Ionides, E. L., D. Nguyen, Y. Atchadé, S. Stoev and A. A. King, "Inference for dynamic and latent variable models via iterated, perturbed bayes maps", Proceedings of the National Academy of Sciences **112**, 3, 719–724 (2015).

Khan, M. U., M. Shahidullah, W. U. Ahmed, D. Purification and M. A. Khan, "The eltor cholera epidemic in dhaka in 1974 and 1975", Bulletin of the World Health Organization **61**, 4, 653–659 (1983).

King, A. A., E. L. Ionides, M. Pascual and M. J. Bourma, "Inapparent infections and cholera dynamics.", Nature **454**, 7206, 877–880 (2007).

King, A. A., D. Nguyen and E. L. Ionides, "Statistical inference for partially observed markov processes via the r package pomp", Journal of Statistical Software **69**, 12, 1–43 (2016).

Koelle, K. and M. Pascual, "Disentangling extrinsic from intrinsic factors in disease dynamics: a nonlinear time series approach with an application to cholera.", The American naturalist **163**, 6, 901–913 (2004).

Lanchier, N., *Stochastic Modelling* (Springer Nature, 2017).

Lange, K., "A gradient algorithm locally equivalent to the em algorithm", Journal of the Royal Statistical Society: Series B (Methodological) **57**, 2, 425–437 (1995).

Lange, K., *Numerical Analysis for Statisticians* (Springer Publishing Company, Incorporated, 2012), 2nd edn.

Lele, S. R., B. Dennis and F. Lutscher, "Data cloning: easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods", Ecology Letters **10**, 7, 551–563 (2007).

Leung, P., C. S. Forbes, G. M. Martin and B. McCabe, "Forecasting observables with particle filters: Any filter will do!", URL `https://arxiv.org/abs/1908.07204` (2019).

Lindsten, F., M. I. Jordan and T. B. Schön, "Particle gibbs with ancestor sampling", Journal of Machine Learning Research **15**, 63, 2145–2184 (2014).

Loader, C. R., "A Log-Linear Model for a Poisson Process Change Point", The Annals of Statistics **20**, 3, 1391 – 1411 (1992).

Lux, T., "Bayesian estimation of agent-based models via adaptive particle markov chain monte carlo", Computational Economics (2021).

Murray, L. M., "Bayesian state-space modelling on high-performance hardware using libbi", (2013).

Owen, J., D. J. Wilkinson and C. S. Gillespie, "Scalable inference for markov processes with intractable likelihoods", Stat Comput **25**, 145–156 (2014).

Pinto, H., S. Howell and D. Paxson, "Modelling the number of customers as a birth and death process", The European Journal of Finance **15**, 2, 105–118 (2009).

Pitt, M. K., R. dos Santos Silva, P. Giordani and R. Kohn, "On some properties of markov chain monte carlo simulation methods based on the particle filter", Journal of Econometrics **171**, 2, 134–151, bayesian Models, Methods and Applications (2012).

Prado, R. and M. West, *Time Series Modeling, Computation and Inference* (Taylor & Francis, CRC Press, Boca Raton, Fl., 2010).

Rabbani, G. and W. Greenough, "Food as a vehicle of transmission of cholera", Journal of Diarrhoeal Diseases Research **17**, 1, 1–9 (1999).

Raftery, A. E. and V. E. Akman, "Bayesian analysis of a Poisson process with a change-point", Biometrika **73**, 1, 85–89 (1986).

Rasmussen, D. A., O. Ratmann and K. Koelle, "Inference for nonlinear epidemiological models using genealogies and time series", PLoS Computational Biology **7**, 8 (2011).

Reuter, G., "Competition processes", In Proceedings of the 4th Berkeley Symposium Mathematical Statistics and Probability, Berkeley, CA, USA **2**, 421–430 (1961).

Robert, C. and G. Casella, *Introducing Monte Carlo Methods with R.* (Springer-Verlag, 2009).

Scarrott, C. and A. MacDonald, "A review of extreme value threshold estimation and uncertainty quantification", REVSTAT - Statistical Journal **10**, 1, 33–60 (2012).

Simons, E., M. Ferrari, J. Fricks, K. Wannemuehler, A. Anand, A. Burton and P. Strebel, "Assessment of the 2010 global measles mortality reduction goal: results from a model of surveillance data.", The Lancet **379**, 9832, 2173–2178 (2012).

Smith, R. L., "Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone", Statistical Science **4**, 4, 367–377 (1989).

Tancredi, A., C. Anderson and O. A., "Accounting for threshold uncertainty in extreme value estimation", Extremes **9**, 87–106 (2006).

Tong, H., "Threshold models in time series analysis—some reflections", Journal of Econometrics **189**, 2, 485–491, frontiers in Time Series and Financial Econometrics (2015).

Tsay, R. S. and R. Chen, *Nonlinear Time Series Analysis* (Wiley Series in Probability and Statistics, 2018).

Wang, C., H. Liu, J.-F. Yao, R. A. Davis and W. K. Li, "Self-excited threshold poisson autoregression", Journal of the American Statistical Association **109**, 506, 777–787 (2014).

West, W. R. and T. R. Ogden, "Continuous-time estimation of a change-point in a poisson process", Journal of Statistical Computation and Simulation **56**, 4, 293–302 (1997).

WHO, "Global framework for immunization monitoring and surveillance : Gfims", (2007).

Wilkinson, D. J., *Stochastic Modelling for Systems Biology* (Boca Raton, Florida: Chapman and Hall/CRC., 2006).

Xu, J., P. Guttorp, M. Kato-Maeda and V. N. Minin, "Likelihood-based inference for discretely observed birth-death-shift processes, with applications to evolution of mobile genetic elements", Biometrics **71**, 1009–1021 (2014).