A sequential Monte Carlo approach to estimate a time varying reproduction number in infectious disease models: the Covid-19 case.

Geir Storvik[1]   Alfonso Diz-Lois Palomares[2]   Solveig Engebretsen[3]
Gunnar Øyvind Isaksson Rø[2]   Kenth Engø-Monsen[4]
Anja Bråthen Kristoffersen [2]   Birgitte Freiesleben de Blasio[2,5]
Arnoldo Frigessi[5,6]

May 3, 2022

[1] *Department of Mathematics, University of Oslo*
[2] *Department of Method Development and Analytics, Norwegian Institute of Public Health*
[3] *Norwegian Computing Center*
[4] *Telenor Research, Fornebu, Norway*
[5] *Oslo Centre for Biostatistics and Epidemiology, University of Oslo*
[6] *Oslo Centre for Biostatistics and Epidemiology, Oslo University Hospital*

Abstract

The Covid-19 pandemic has required most countries to implement complex sequences of non-pharmaceutical interventions, with the aim of controlling the transmission of the virus in the population. To be able to take rapid decisions, a detailed understanding of the current situation is necessary. Estimates of time-varying, instantaneous reproduction numbers represent a way to quantify the viral transmission in real time. They are often defined through a mathematical compartmental model of the epidemic, like a stochastic SEIR model, whose parameters must be estimated from multiple time series of epidemiological data. Because of very high dimensional parameter spaces (partly due to the stochasticity in the spread models) and incomplete and delayed data, inference is very challenging. We propose a state space formalisation of the model and a sequential Monte Carlo approach which allow to estimate a daily-varying reproduction number for the Covid-19 epidemic in Norway with sufficient precision, on the basis of daily hospitalisation and positive test incidences. The method was in regular use in Norway during the pandemics and appears to be a powerful instrument for epidemic monitoring and management.

Keywords: Covid-19, SEIR model, reproduction number, Sequential Monte Carlo, infectious disease, epidemic modelling, Bayesian estimation, multiple data sources, real time situation awareness.

1   Introduction

We propose a dynamic approach for the estimation of a time-varying or instantaneous reproduction number for a mathematical infectious disease spread model. We apply our method to the Covid-19 pandemic in Norway. Like in most other countries, the pandemic has been tackled with a combination of non-pharmaceutical interventions, from social distancing to partial lock-down, imposed or advised at various time points. Various viral variants with different characteristics

have been competing in the population. Vaccination has also been gradually introduced. As a consequence of these changes, and of varying human behaviour, which can emerge both abruptly and smoothly, the reproduction number varies. Instantaneous estimates of reproduction numbers are useful for situational awareness. Being able to estimate such changes rapidly is important in guiding decision makers in future policy planning.

A reproduction number is precisely defined within a mathematical model of transmission. A large class of models, which has been shown to be very appropriate, is the so called SEIR model (S=susceptible, E=exposed, I=infected, R=recovered). In particular, *stochastic* compartmental models are preferable when there are relatively low number of infections (Keeling and Rohani, 2011, chapter 6). SEIR models are parametrised so that a meaningful estimate of the reproduction number (basic and effective) can be derived. Such compartmental models are based on many latent variables, in particular the number of individuals in each compartment within each region at every time point, and depend on many epidemiological parameters, including the transmission strengths which are key components of reproduction numbers. All unknowns must be estimated from data, which in a realistic situation are scarce and incomplete. For this reason, inference is in general very difficult, because of a high dimensional parameter space, rather flat likelihoods or posterior distributions and, often, weak identifiability (De Angelis et al., 2015).

Moreover, data which carry information about the transmissibility of the virus appear with an inevitable delay. In this paper we use two data sources, the daily number of patients hospitalised because of Covid-19 and the daily number of positive laboratory-confirmed RNA test cases. Both these data sources carry information about the transmission of the virus in the society, but with a random time delay from transmission. Therefore, the uncertainty of the estimates of a daily reproduction number increases in the last period of data. All this is particularly challenging during an emerging epidemic. For these reasons, instantaneous reproduction numbers are rarely assumed in SEIR models. In this paper we propose and test a sequential Monte Carlo (SMC) approach to inference which allows the efficient estimation of instantaneous reproduction numbers for the Covid-19 pandemic from real data.

While our SMC approach is generic, we implement it on top of a stochastic SEIR model which we developed for the Covid-19 epidemic (Engebretsen et al., 2021) and which was in regular use by the Norwegian health authorities. This model assumes a spatial scale resolved on county level, and uses mobile phone mobility data for the geographical spread of the virus in temporal steps of six hours. In Engebretsen et al. (2021), the transmissibility parameter, which represents the probability of transmission upon a contact times the contact rate in the population, is assumed to be constant in time, and is only changed at designed change-points. Inference in Engebretsen et al. (2021) is performed by a version of Approximate Bayesian Computation (ABC). MCMC convergence was essentially impossible to reach, because of the difficulty to design parameter perturbations which would lead, through the stochastic SEIR, to minor and controlled changes of the posterior distribution. It is very difficult to use ABC when the number of parameters is large, as is the case when including daily varying reproduction numbers.

In this paper we propose to perform Bayesian inference in combination with SMC. Static parameters related to the dynamic process for the reproduction number are estimated sequentially through SMC using sufficient statistics (Fearnhead, 2002; Storvik, 2002). Application of SMC methods is challenging because the latent processes are of high dimensions, the SEIR model is only available as a computer algorithm, and data are very limited. A further very important practical aspect when working in real time during a pandemic, comes from the continuous need to improve the model, to change epidemiological assumptions, to use different data registries, and to improve the computational efficiency of algorithms. For these reasons, it has been very important to develop an SMC which is very flexible, so that changes in model specifications are easy to implement. This paper shows how a careful design of combination of model and algorithm reaches these aims and leads to a good fit to both data sources. We produce an analysis of the Norwegian pandemic, quantifying how interventions impacted the transmissibility and showing how our estimated instantaneous reproduction numbers were capturing changes rapidly enough. Based on the dynamic model, we perform future predictions of the situation for the next three weeks. We discuss the quantification of uncertainty in forecasts, which is of paramount importance

for decision making.

There is an important literature on time varying reproduction numbers applied in various epidemics, see for example Cauchemez et al. (2006); Viboud et al. (2018). A Bayesian framework for estimating time-varying reproduction numbers was proposed by Cauchemez et al. (2006), and applied to the SARS epidemic. In Cori et al. (2013) a time varying reproduction number is defined as the ratio of the number $C(t)$ of infected in day t over $\sum_k w(k)C(T-k)$, where $w(\cdot)$ is the distribution of the generation time of Covid-19, which is often set to be distributed with a mean and a standard deviation estimated from specific studies (Ferretti et al., 2020). There is a very successful R-package implementing this Bayesian method, called `EpiEstim` (Cori et al., 2021). We include a comparison between `EpiEstim` and our approach based on only one of the two data sources.

A series of interesting papers has been recently presented in the Royal Statistical Society Special Topic Meeting on Covid-19 transmission (RSS, 2021). Pellis et al. (2022) give a thorough discussion of the reproduction number, its model-dependent definition, its estimation and the challenges in communicating it to the public. A model-agnostic alternative to the reproduction number is the instantaneous epidemic growth rate, and the two measures are compared in Parag et al. (2021). Mishra et al. (2020) extend in various ways the renewal model in Flaxman et al. (2020) and assume a random walk model for a weekly constant reproduction number. In order to study the effect of school closures, Dunbar and Held (2021) fit a multivariate time-series model to case incidence data and perform an interesting counterfactual analysis. Although in the present paper we concentrate inference about the reproduction numbers, we emphasise that several other measures can be extracted from our model, as shown in the weekly reports published by the Norwegian Institute of Public Health during the pandemic (NIPH, 2022).

Several papers have applied MCMC algorithms to estimate parameters in compartmental models (e.g Gibson and Renshaw, 1998; O'Neill and Roberts, 1999; O'Neill et al., 2000) as well as alternative models (Teh et al., 2021). The recent paper Birrell et al. (2020) studies various SMC approaches in a SEIR model for influenza. It demonstrates the superiority of the SMC methodology compared to MCMC for such dynamical models. The paper is also useful as a general reference to SMC in epidemic modelling. Our work shares many similarities to this approach, including the use of several sources of data. A difference is our use of a dynamic model for the reproduction numbers. Also, a stochastic delay between infection and observation time is included in our setting. Another general inference framework is implemented in the R-package `pomp` (King et al., 2016). The package contains multiple different implementations of estimation procedures, including Sequential Monte Carlo, for inference for partially observed Markov process models. There are several examples of applications of the package to epidemic inference, see for example (King et al., 2016; Stocks et al., 2020; Blackwood et al., 2013).

The outline of the rest of the paper is as follows. In Section 2 the context and the data are described. In section 3 we specify the full model, formulated as a state space model. In Section 4 we discuss how SMC algorithms can be applied in inference, including estimation of several static parameters. A simulation study and experimental results are reported in Section 5 specifically for the Norwegian Covid-19 pandemic. Additional results, including sensitivity analyses, are collected in the supplementary material. We conclude the paper with a summary and discussion. In the supplementary material details about experimental settings and algorithmic specifications are given. Data and codes are available on our GitHub repository.

2 Context and data

We start by setting the scene of the inferential task. The core is an existing model of the epidemic which has as input a set of parameters and variables, including daily reproduction numbers, and as output a series of time series of infection incidence. In our case, the model is a stochastic compartmental SEIR-type model that produces numbers of susceptible, exposed, pre-symptomatic, symptomatic and asymptomatic infectious and recovered at every time point. We also keep track of the disease incidence. We use two data time series to inform the SEIR model: the daily number of new hospital admissions of Covid-19 patients, and the daily number of laboratory-confirmed

positive PCR tested cases. In order to exploit these data, we furthermore model the process of hospitalisation and testing of the SEIR output, in particular of the daily incidence of infected. The inferential task is to estimate the input parameters.

The hospitalisation data contain admission to all hospitals in Norway of all patients who are admitted with Covid-19 as the main cause. Admission on a certain day informs us of a transmission event that has occurred some days before. This time gap can differ between individuals. We make several assumptions on various time lags, as specified in supplementary material. For example, the number of days between symptom onset and admission to hospital is estimated to be negative binomial distributed with parameters estimated in a separate study of the Norwegian Covid-19 registry (Whittaker et al., 2021). On average, for a patient being hospitalised, the time gap between infection and hospitalisation is estimated to be approximately 14 days. For concreteness, in this paper we assumed the distributions of the various time lags to be given.

The second data set is the time series of daily number of positive PCR tests. Again, there is a time gap between onset of symptoms and testing, which we estimate to be randomly distributed with mean about 4 days. It is important for inference that the two data sets are as coherent as possible.

We use two additional data sets, which enter the SEIR model as input variables: the daily number of positive cases who have been tested in Norway but infected outside of Norway, so called imported cases; and the total number of PCR tests made in Norway, as a surrogate of the effort made to detect positive cases.

We start with the population of Norway, distributed according to the national census in the eleven counties (see www.ssb.no/en/statbank/table/12871/). Like in Engebretsen et al. (2021), we seed our model continuously with positive cases imported from abroad on the day of recorded symptom onset or, if not available, when detected by testing. Imported cases that are hospitalised are not counted in the time series of hospitalisations, because they do not inform the model about the transmissibility of the virus in Norway. Because not all imported cases are likely to be discovered, we assume that each imported case stands for an unknown number of further undetected imported cases. We model this latent import with an additional Poisson distributed number of cases per observed imported case, with Poisson mean estimated from the data during calibration. We call this mean the amplification factor.

A final aspect, which is not central in this paper but that we mention for completeness, is that we use a geographical SEIR model on county (regional) level, so that the various compartments are geographically defined. Individuals are moved at random between the eleven counties of Norway using a mobility matrix, which is obtained every six hours from the movements of mobile phones in Norway, as explained in Engebretsen et al. (2021). In the present paper, all parameters in the model are however shared between counties. Even if hospitalisation and test data are available at county level, in this paper we use only nationally aggregated data, because of the heterogeneity in the population size among the regions. A very important further aspect is the need to obtain inferential results as rapidly as possible, in at most a few hours, so to be able to publish results quickly just after release of the data update.

3 Model

Let $\boldsymbol{y}_t$ be the vector of hospitalisation and test data on day t. Let $\boldsymbol{s}_t$ be the output vector of compartmental variables in the SEIR model at time t, for example the number of individuals in each county who are infected and symptomatic. Here we consider the model generically as an algorithm which outputs the compartmental variables $\boldsymbol{s}_t$ at each time point t. R_t is the unknown reproduction number at time t. We consider the following state space model:

$$R_t \sim P_R(R_{t-1}; dR_t); \qquad \text{for the reproduction number,} \qquad (1a)$$

$$\boldsymbol{s}_t \sim P_s(R_t, \boldsymbol{s}_{t-1}; d\boldsymbol{s}_t); \qquad \text{representing the SEIR process,} \qquad (1b)$$

$$\boldsymbol{y}_t \sim H_{y|s}(\boldsymbol{s}_{1:t}; d\boldsymbol{y}_t); \qquad \text{for the hospital and test data.} \qquad (1c)$$
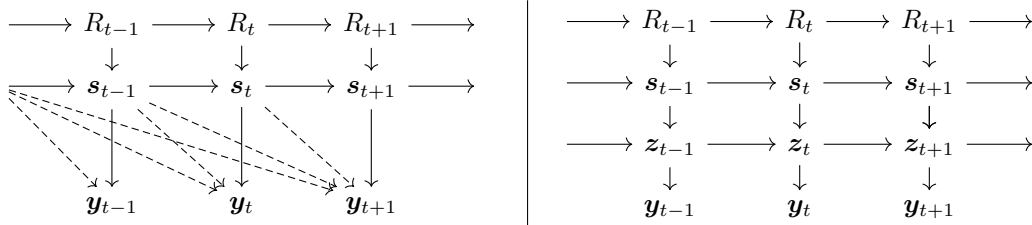
Figure 1: Graphical representation of model (1) (left) and its reformulation (right). The hyper-parameters $\boldsymbol{\theta}$ are not included and can influence all conditional distributions. The dashed arrows illustrate the dependence due to the random delays between infections and testing and possibly hospitalisation.

To simplify notation, we do not include the dependence of the models on the set of static parameters $\boldsymbol{\theta}$. The distribution P_R needs to be available analytically and easy to sample from. The distribution P_s is assumed to be only available through a computer algorithm and we are only able to simulate from this distribution. In certain situations, this distribution can be available as a huge and complex Markov process. However this is not often the case, for example because of the complexity of the code or because of the lack of availability of sensitive data, like the mobility matrices in our case. The dimension of $\boldsymbol{s}_t$ is large while R_t is low-dimensional. In this work we consider a common scalar R_t for all counties. Note that the data $\boldsymbol{y}_t$ depend on the whole history $\boldsymbol{s}_{1:t}$ making $\boldsymbol{y}_t$ only weakly informative about $(R_t, \boldsymbol{s}_t)$. This is due to the fact that there is a random delay from transmission to being tested and possibly hospitalised. A graphical representation of the model is given in the left panel of Figure 1. Our aim is to construct an efficient SMC method for the computation of $p(R_t, \boldsymbol{s}_t | \boldsymbol{y}_{1:T})$ and we are interested in estimating the current status $(t = T)$, in smoothing $(t < T)$ and in forecasting $t > T$.

The stochastic process $\{R_t\}$ is assumed to be Markov. We suggest three alternative prior models for P_R. Let $\varepsilon_t \sim N(0, \sigma_R^2)$ be independent from all other variables. The first model is a random walk on log-scale, the second model extends this to an AR(1) structure. In the third model R_t is assumed to be piece-wise constant, with jumps occurring at random.

$$R_t = e^{\log(R_{t-1}) + \varepsilon_t} \qquad\qquad \text{RW, model 1}$$

$$R_t = e^{\mu + a(\log(R_{t-1}) - \mu) + \varepsilon_t} \qquad\qquad \text{AR, model 2}$$

$$R_t = \begin{cases} R_{t-1} & \text{with probability } 1 - \phi_c \\ R_{t-1} e^{\varepsilon_t} & \text{with probability } \phi_c. \end{cases} \qquad \text{PC, model 3}$$

For all models, we assume a constant start: Until a given time point D, R_t is constant and equal to R_0. The date D is set to the day when the social-distancing implementation started. Before D the reproduction number was not likely to change significantly. In this paper we set D equal to March 8, 2020, when teleworking started in many companies and universities in Norway. The Norwegian government announced the first package of interventions on March 12 2020. The models above describe the dynamics after D. All the static parameters in the three models must be estimated.

The proposed method is not dependent on a specific model (1b), and would work with any epidemic model. In this work we use the particular SEIR model described in Engebretsen et al. (2019, 2020), and whose algorithm is available though the `spread` package (Engebretsen et al., 2020).

In particular, this SEIR model has six compartments in each region (county) i: susceptible (S), exposed and not infectious (E_1), presymptomatic and infectious (E_2), infectious symptomatic (I), infectious asymptomatic (I_a), and recovered (R). The dynamics is described by the following

equations

$$
\begin{aligned}
S^i_{t+\delta_t} &= S^i_t - X^i_{1,t} & X^i_{1,t} &\sim \text{Binom}(S^i_t, \beta_t(I^i_t + r_{I_a}I^i_{a,t} + r_{E_2}E^i_{2,t})\delta_t) \\
E^i_{1,t+\delta_t} &= E^i_{1,t} + X^i_{1,t} - X^i_{2,t} & X^i_{2,t} &\sim \text{Binom}(E^i_{1,t}, \lambda_1\delta_t)/N^i_t \\
E^i_{2,t+\delta_t} &= E^i_{2,t} + X^i_{3,t} - X^i_{4,t} & X^i_{3,t} &\sim \text{Binom}(X^i_{2,t}, (1-p_a)\delta_t) \\
& & X^i_{4,t} &\sim \text{Binom}(E^i_{2,t}, \lambda_2\delta_t) \\
I^i_{t+\delta_t} &= I^i_t + X^i_{4,t} - X^i_{5,t} & X^i_{5,t} &\sim \text{Binom}(I^i_t, \gamma\delta_t) \\
I^i_{a,t+\delta_t} &= I^i_{a,t} + X^i_{2,t} - X^i_{3,t} - X^i_{6,t} & X^i_{6,t} &\sim \text{Binom}(I^i_{a,t}, \gamma\delta_t),
\end{aligned}
\tag{2}
$$

where δ_t is 6 hours. We assume random mixing in each county in each 6 hour period, and move individuals between counties at the end of each such period according to a mobility matrix. In Engebretsen et al. (2021) mobile phone data are used to estimate such mobility matrices. These matrices report the number of individuals moving from county A to county B during each period, which we pick at random among those currently present in A, but favour the residents of B to return to B, to capture commuting. This rule makes the computational complexity of the SEIR model quadratic in the number of counties, due to the need for storage of both current visited location and residence of individuals. The reproduction number R_t is related to β_t through the equation

$$
R_t = \beta_t \left(\frac{1 - p_a + p_a r_{I_a}}{\gamma} + \frac{(1-p_a)r_{E_2}}{\lambda_2} \right).
$$

Finally we describe the likelihood model for the data (1c). A main difficulty is the link between the latent process $\{s_t\}$ and the observation process $\{y_t\}$ because of the unknown stochastic delays between infection and observation time, making the computation of $H_{y|s}(s_{1:t}; dy_t)$ hard. We introduce an auxiliary process $z_t = (z^H_{t,0:L^H}, z^T_{t,0:L^T})$ with two components, one dedicated to the hospitalisation and the other to the test data. The auxiliary variable $z^H_{t,v}$ is defined as the number of individuals who are infected at time t and hospitalised v days later. The time lag v is assumed to vary in $\{0, ..., L^H\}$, for some appropriate L^H. Similarly $z^T_{t,v}$ is the number of individuals who are infected at time t and tested positive v days later. We rewrite (1c) as

$$
\begin{aligned}
z_t &\sim G(z_{t-1}, s_t; dz_t), & \text{(3a)} \\
y_t &\sim H_{y|z}(z_{1:t}; dy_t), & \text{(3b)}
\end{aligned}
$$

where $G(z_{t-1}, s_t; dz_t)$ is a Markov transition distribution assumed to be easy to simulate from. Now $H_{y|z}(z_{1:t}; dy_t)$ is easy to compute. In more details, define y^H_t to be the number of daily Covid-19 admissions to hospital. An individual who is infected at time t is hospitalised with probability p^H_u at time $t + u$. The time lag u from infection until hospitalisation is assumed to follow a discrete distribution on the integers $\{0, 1, ..., L^H\}$. Let ρ^H_u be the probability of delay u. To structure the model further, we attach to each infected individual its potential time lag until hospitalisation. So $z^H_{t,u}$ is the number of individuals infected at time t and who could possibly be hospitalised u time-units later. Let I_t be the total number of infected individuals at time t, which is available through the SEIR model as a component of s_t. We can formulate the model as

$$
\begin{aligned}
z^H_{t,0} | I_t &\sim \text{Binomial}(I_t, \rho^H_0); \\
z^H_{t,u} | I_t, z^H_{t,0:u-1} &\sim \text{Binomial}\left(I_t - \sum_{v=0}^{u-1} z^H_{t,v}, \frac{\rho^H_u}{1 - \sum_{v=0}^{u-1} \rho^H_v} \right) \quad u = 1, ..., L^H; \\
y^H_t &\sim \text{Beta-Binomial}(\sum_{u=0}^{L} z^H_{t-u,u}, \alpha^H, \beta^H_t),
\end{aligned}
\tag{4}
$$

where we here consider a Beta-Binomial distribution for a patient being hospitalised to take into account variability in hospitalisation between regions. The β^H_t parameter is specified indirectly

Algorithm 1 Auxiliary SMC with resampling at each time step. Operations involving index b must be performed for $b = 1, ..., B$. Here P_t^x denotes the transition distribution for $\boldsymbol{x}_t$, while Q_t is the proposal distribution for $\boldsymbol{x}_t$. The indices $A_t^{1:B}$ defines the ancestral particles at time t after resampling.

1: $\boldsymbol{x}_0^b \sim Q_0(d\boldsymbol{x}_0)$ ▷ Proposal at time 0

2: $w_0^b = \frac{P_0(d\boldsymbol{x}_0^b)H(\boldsymbol{y}_t|\boldsymbol{x}_0^b)}{Q_0(d\boldsymbol{x}_0^b)}, \quad W_0^b = w_0^b / \sum_{m=1}^B w_0^m$ ▷ Calculating weights

3: $\ell_0^B = \frac{1}{B} \sum_{b=1}^B w_0^b$ ▷ Estimate of $p(\boldsymbol{y}_0)$

4: **for** $t = 1$ to T **do**

5:      $A_t^{1:B} = \text{resample}(W_{t-1}^{1:B})$ ▷ Resampling

6:      $\boldsymbol{x}_t^b \sim Q_t(\boldsymbol{x}_{t-1}^{A_t^b}, d\boldsymbol{x}_t)$ ▷ Proposal at time t

7:      $w_t^b = w_{t-1}^b \frac{P_t(\boldsymbol{x}_{t-1}^{A_t^b}, d\boldsymbol{x}_t^b)H(\boldsymbol{y}_t|\boldsymbol{x}_t^b)}{Q_t(\boldsymbol{x}_{t-1}^{A_t^b}, d\boldsymbol{x}_t^b)}, \quad W_t^b = w_t^b / \sum_{m=1}^B w_t^m$ ▷ Calculating weights

8:      $\ell_t^B = \frac{1}{B} \sum_{b=1}^B w_t^b$ ▷ Estimate of $p(\boldsymbol{y}_t|\boldsymbol{y}_{t-1})$

9: **end for**

---

through a time-varying probability p_t^H such that $\beta_t^H = \alpha(1 - p_t^H)/p_t^H$ where p_t^H is predefined using the age-structure of the individuals having tested positive. By storing and sequentially updating the quantities $\sum_{u=0}^L z_{t-u,u}^H$ as well, we obtain a first order Markovian state space structure as illustrated in the right panel of Figure 1.

Regarding the test data, as in Engebretsen et al. (2021) we consider the probability that an infected case is tested by means of a PCR test. We ignore that tests can lead to false positive responses. The logit of this probability is assumed to be linear in the total number of daily tests N_t^T in day t in addition to a time independent intercept. We write the detection probability ρ_t^T at time t as

$$\rho_t^T = \frac{\exp(\pi_0 + \pi_1 N_t^T)}{1 + \exp(\pi_0 + \pi_1 N_t^T)}. \tag{5}$$

The time lag between infection and testing is assumed to follow a discrete distribution on $\{0, 1, ..., L^T\}$ for an appropriate L^T. The approach for handling this delay is exactly as for the hospital incidence, with p_t^T now playing the role of p_t^H. We introduce a new set of auxiliary variables for the test data $\{z_{t,u}^T\}$, similarly to the ones introduced for the hospitalisation data. Defining $\boldsymbol{z}_t = (\boldsymbol{z}_{t,0:L^H}^H, \boldsymbol{z}_{t,0:L^T}^T)$, we are within the model formulation (3).

4 Sequential Monte Carlo

Let $\boldsymbol{x}_t = (R_t, \boldsymbol{s}_t, \boldsymbol{z}_t)$. Our aim is to perform inference on the whole set of latent variables $\boldsymbol{x}_{1:t} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_t)$ as well as on static hyper-parameters $\boldsymbol{\theta}$ at each time-point t, by means of the posterior distribution $p(\boldsymbol{x}_{1:t}, \boldsymbol{\theta}|\boldsymbol{y}_{1:t})$. A description of the SMC algorithm with resampling at each step is given in Algorithm 1. See also Chopin and Papaspiliopoulos (2020), which contains more general algorithms. Although in this paper we focus on inference for R_t, also $\boldsymbol{s}_t$ will be of interest. In our setting, the main computational burden is the sampling from $P_s(R_t, \boldsymbol{s}_{t-1}; d\boldsymbol{s}_t)$ which has been parallelised in our implementation. For resampling, residual resampling (Liu and Chen, 1998) has been applied. However, the resampling step is both hard to parallelise and requires message passing, resulting in that a too high number of cores can decrease performance.

The SMC algorithm is by design sequential so that by storing values of $\boldsymbol{x}_t$ obtained at the previous day, updates can easily be performed as new data arrive. A main challenge here is that the state $\boldsymbol{x}_t$ at time t heavily depends on *future* observations $\boldsymbol{y}_{t+h}$ because of the delay in hospitalisation. Although the reformulated model reduces immediate dependence, there are still strong correlations backwards, as illustrated in Figure 1. There are clearly possibilities to develop more efficient proposal distributions, despite the availability of the SEIR model only as a computer algorithm. Because the main purpose of this paper is to investigate the utility of SMC methods for

the estimation of daily varying reproduction numbers, we use only simple bootstrapping proposals. We do however take into account the delay aspect by using fixed lag smoothing, using data ahead of current time, that is

$$\widehat{R}_{t|t+l_t} = E[R_t|\boldsymbol{y}_{1:t+l_t}] \approx \frac{1}{B} \sum_{b=1}^{B} R_t^b$$

where $\{R_t^b\}$ are simulations of R_t based on data up to time point $t + l_t$. Fixed lag smoothing with a lag of $l_t = 24$ days has been used in the Covid-19 runs in this paper. At the end of the time-series, $l_t = \min\{T - t, 24\}$ is used. The estimates on the last days will be more uncertain.

4.1 Parameter estimation

Algorithm 1 assumes that the parameters $\boldsymbol{\theta}$ are known. Now we describe Bayesian inference for some of the static parameters. We denote by $\boldsymbol{\theta}_R$ the set of parameters in the model for $\{R_t\}$, $\boldsymbol{\theta}_s$ the ones in the $\{\boldsymbol{s}_t\}$ process, and $\boldsymbol{\theta}_y$ the parameters appearing in the data model. As mentioned, some of these parameters are fixed based on other data sources, and here for simplicity we do not propagate their uncertainty. For other parameters, sequential updates of their estimates are desirable. In principle, all parameters $\boldsymbol{\theta}$ could be included as part of the state vector $\boldsymbol{x}_t$ where the propagation of these static components just keeps them fixed. However, repeated resampling will quickly give degenerate samples for these parameters.

A review of parameter estimation in SMC is given in Kantas et al. (2015). Off-line methods such as Particle MCMC (PMCMC, Andrieu et al., 2010) have proven to be very effective in many applications, but require repeated runs of the SMC routine. Although much smaller number of particles can be applied in such settings, some experiments with our models indicate that at least 250 particles are necessary, in which case one run uses about 10 minutes using 4 cores on Linux server and more cores did not help much in this case. Some experiments with an implementation of the Particle Metropolis-Hastings algorithm, based on the pseudo-marginal method by Andrieu et al. (2009), is reported in supplementary section D. We will however focus on online methods for parameter estimation here.

In cases where sufficient statistics $\boldsymbol{v}_t(\boldsymbol{x}_{1:t})$ for the parameters are available, the SMC algorithm can be easily updated to target $p(\boldsymbol{x}_{1:t}, \boldsymbol{v}_t(\boldsymbol{x}_{1:t})|\boldsymbol{y}_{1:t})$ instead (Fearnhead, 2002; Storvik, 2002). This is the case for the $\boldsymbol{\theta}_R$ parameters. Simulations of $\boldsymbol{\theta}_R$ at each time point can then be obtained from $p(\boldsymbol{\theta}_R|\boldsymbol{v}_t(\boldsymbol{x}_{1:t}))$ which then again can be used to obtain new samples of $\boldsymbol{x}_{t+1}$ (step 6 in Algorithm 1). A crucial step is that $\boldsymbol{v}_t(\boldsymbol{x}_{1:t})$ can be recursively updated. Section A in the supplementary material gives details on how $\boldsymbol{\theta}_R$ can be updated by this approach. Note that these methods can suffer from the degeneracy problem (Andrieu et al., 2005). In supplementary section D we validate the parameter estimates obtained by this procedure both through comparisons between different runs and by using the samples obtained by the Particle Metropolis-Hastings algorithm.

5 Results

5.1 Norwegian Covid-19 data

For the analysis of the Norwegian Covid-19 case, we used the 11 counties as our spatial scale. Mobility data and imported cases were used at regional scale while hospital incidence data and test data were used at national scale. Also the $\{R_t\}$ prior process was assumed to be common for all regions. The hospital incidence data are from the Norwegian national Beredt-C19 registry and the Norwegian Intensive Care and Pandemic Registry, and the test data are from the Norwegian Surveillance System for Communicable Diseases (MSIS registry). The reproduction number is assumed constant until March 7 2020. The results are based on data up to July 1, 2021 with test data included only from August 1, 2020, when testing capacity in Norway was scaled as needed, and after which testing criteria had become rather stable. The number of parameters is large: the dimension of $\boldsymbol{s}_t$ is 3157. In addition we have the reproduction numbers R_t and the auxiliary variables $\boldsymbol{z}_t$'s. The prior distributions assumed for the parameters involved in the dynamics of R_t

| | Model | | |
Prior	RW	AR	CP
$a \sim \mathrm{N}(0.5, 0.25)$		0.563 (0.430,0.659)	
$\frac{1}{\sigma^2} \sim \mathrm{Gamma}(2.4, 0.28)$	0.176 (0.157,0.202)	0.340 (0.300,0.477)	0.574 (0.485,0.678)
$\phi \sim \mathrm{Beta}(1, 9)$			0.158 (0.119, 0.198)
MLIK	-3028.20	-3008.68	-3020.69
MLIK-no test	-1252.15	-1221.09	-1230.71

Table 1: Posterior medians and 95% credibility intervals (in parentheses) for hyper-parameters of the three models of the R_t dynamics. Although the prior for σ is defined through the precision, the numbers are for σ itself. The two last rows of the table give the value of the marginal log-likelihoods based on both data sources (MLIK) and on hospitalisation data only (MLIK-no test).

are given in the first column of Table 1. When running the model, we did not estimate the SEIR parameters in equation (2), nor the parameters π_0 and π_1 in (5), because they were estimated separately as described in Engebretsen et al. (2021). Also the parameters in equation (4) were pre-estimated through other data sources. All further details on the model are given in supplementary material section B. Each run is based on $B = 20\,000$ particles. One run of the 500 days considered here, using a linux server with 128 cores, took approximately 5 hours, which is appropriate for practical real-time purposes. Figure 2 shows our estimates of R_t, with a 7-precedent-days moving-average smoothing (daily estimates are given in the supplementary material Figure S3), using the three considered models for the reproduction number. Posterior medians and symmetric 95% credibility intervals of the estimated parameters are reported in the second column of Table 1. The autoregressive model for R_t was simplified by fixing $\mu = 0$: when we estimated μ, we obtained an estimate very close to zero, but we also experienced some difficulties in the estimation of σ^2. In addition, the marginal log-likelihood (MLIK) was slightly higher. We therefore opted for the simpler autoregressive model.

In Figure 2 we see that the different models lead to rather similar estimates for the time dynamics of $\{R_t\}$. The autoregressive model (central panel) seems to have more uncertainty. On the other hand, this model is to be the preferred in terms of MLIK (Table 1, last rows). Note that all models capture quickly the dramatic reduction in transmission in mid March 2020 when Norway had the first lockdown, which was implemented between March 9 and March 14 2020. In the second half of April 2020, society was re-opened, including schools and kindergarten, and the 2-meters distance rule was reduced to 1-meter: $\{R_t\}$ appears to increase to around 1. A new peak appears in the end of July and beginning of August 2020, in correspondence with the end of the Norwegian vacations, the returning of Norwegians from abroad and the arrival in Norway of tourists. In particular there have been several isolated clusters, for example in two cruise ships and the return of students to university campuses. These clusters were well controlled by contract tracing, and the reproduction number dropped rapidly below 1. The autumn 2020 was also characterised by a series of larger local outbreaks, which were rapidly controlled. The growth of R_t starting in October 2020 was also due to local outbreaks, and the incoming winter-season which affects the viral transmission. The number of cases was so large that contact tracing became less effective. The Norwegian government imposed a second national intervention, first in the end of October 2020 and then again in early November 2020. This reduced the reproduction number again to below 1, in two pushes, where we can see that the second strengthening of the intervention was in fact needed for this purpose. We can see that the autumn 2020 interventions allowed the reproduction number to fall from approximately 1.5 to below 1 in about three weeks. Interestingly, we see a peak of R_t just around Christmas 2020, in connection with vacation travel and intensive shopping. January 2021 marks the arrival of the alpha variant of the virus, which was more transmissible and which increased the risk of hospitalisation. The alpha variant was predominant in Norway at the end of March 2021, and we see R_t just below 1.5 again. This increase happened despite at the end of January 2021 Norway introduced the strictest lockdown rules during the whole pandemic so far, including essential closure of all borders, and vaccination
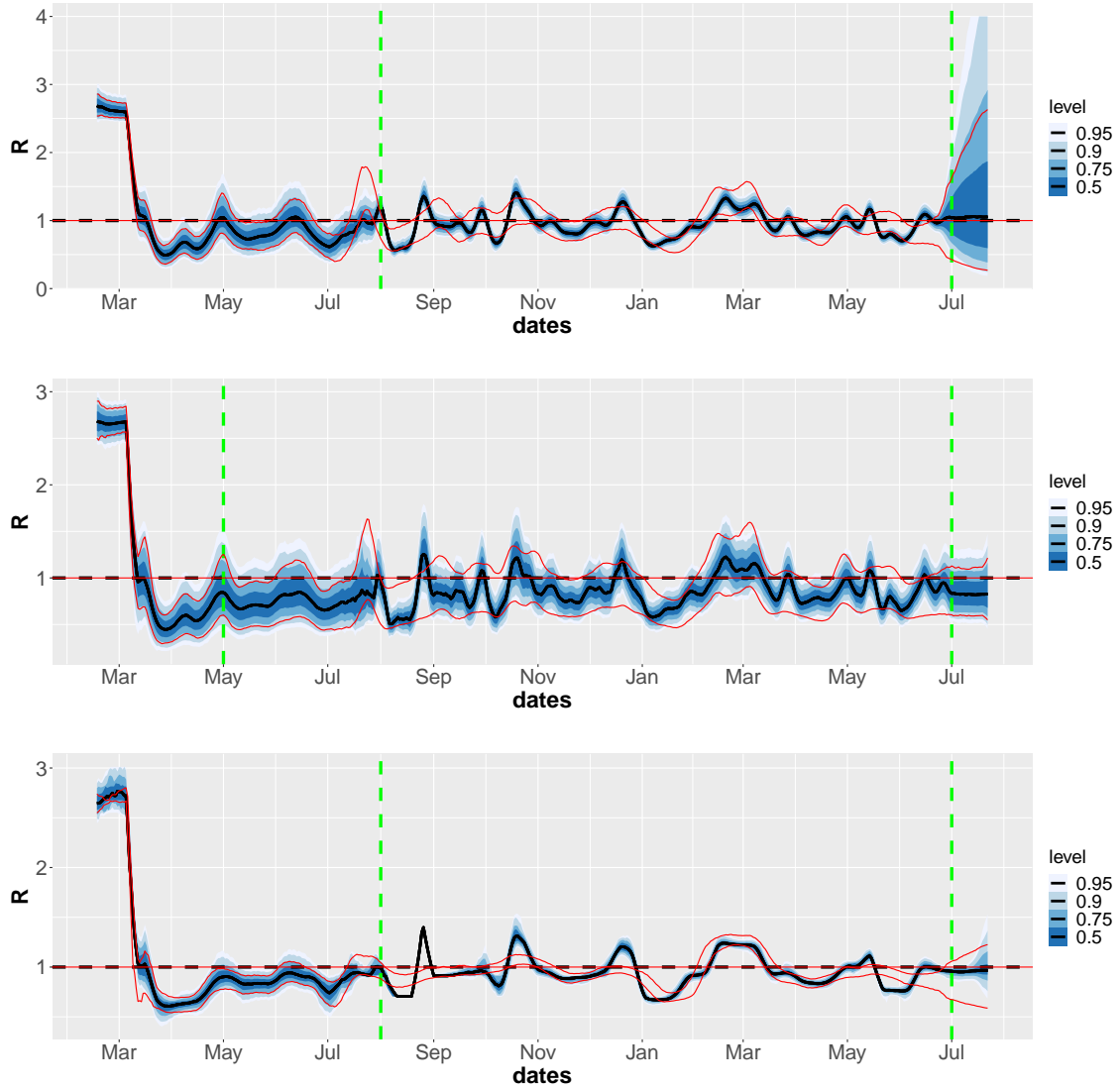
Figure 2: Estimates of the weekly averaged instantaneous reproduction number $\frac{1}{7}\sum_{u=0}^{6} R_{t-u}$ based on the random walk (RW, upper), autoregressive (AR, middle) and piece-wise constant model (PC, lower). We used hospital data from March 8 2020 to July 1 2021, test data from August 1 2020 to July 1 2021. We used $B = 20\,000$ particles. Corresponding not-averaged daily estimates are given in Figure C.2. The first vertical green line corresponds to the date at which test data is included in the analysis, the second green vertical line corresponds to the last date with observations, after that Bayesian prediction is performed. Blue shadow bands indicate posterior uncertainty quantiles. The red curves correspond to 50% median-centred credibility intervals only based on hospitalisation data.

of the elderly started. During March and April 2021, R_t started to decrease again, to remain below or around 1 until the middle of May 2021. Governmental interventions were reduced from May 2021, and the reproduction number stabilised around 1. At the end June 2021, approximately 50% of the adult Norwegian population had been vaccinated at least once, and approximately 30% twice. The effective reproduction number R_t reflects the effect of vaccination which is included in the SEIR model. It is remarkable how the estimated reproduction number quantify the history of the epidemic so precisely. Another feature is the cyclic behaviour of R_t, with a drop following an increase. The AR prior model on log scale also attracts towards 1. In Norway this is expected because of the rapid intervention strategy of the government (named "control") whenever R_t was growing rapidly above 1, and the rapid reopening when the epidemic appeared under control. Local outbreaks were frequent, also visible in the raw data, but they were rapidly controlled by appropriate contact tracing and other successful local interventions.

The red lines in Figure 2 give 50% centred credibility intervals based only on hospitalisation data. We investigate in this way the value of the test data as a second source of information. Because we used test data only from August 1 2020, the estimates are essentially identical until then. After August 1 2020, the estimates based only on hospital data are smoother, indicating that the test data contain information about transmission (and its change) that is not transferred to the hospitalisation. One reason for this is that the younger generations have been infected in the autumn more than the elderly ones, who are most at risk for hospitalisation. The test data also contribute to a more precise estimate of the daily prevalence of infected in Norway. We also observe some misalignment in time between the two estimates, probably because the time delays were not stationary, while we assumed them constant during the whole epidemic.

The last green horizontal line in Figure 2 corresponds to the last day with data, July 1, 2021, after which we perform a three weeks prediction, by simply running the model forward in time. The predictions for the three models have similar means but differ in uncertainty. Both the random walk and the piece-wise constant model are non-stationary, which explains the high increase in uncertainty after the last observation point. On the other hand, the autoregressive model shows a more stable prediction performance, as expected. Figure 3 gives predicted number of new daily infected cases. Before the last green dashed line these credibility intervals give predictions based on the observed data, after this line, the predictions are obtained by running the SEIR model forward three weeks in time using the predicted reproduction numbers. When predicting forward in time, mobility matrices, imported cases and total number of tests are needed as input. We here re-use the values in the previous 21 days in the forecasts. We see the three waves which hit Norway in March/April 2020, from November 2020 to January 2021 and in March 2021. The number of cases is estimated around 1000 per day in the peaks (1300 in the third wave). The number of cases would in fact grow during July 2021, as here correctly predicted.

Table 1 provides the estimates of the parameters in the three dynamic models of $\{R_t\}$. How these estimates are learned over time is shown in Figures C.3-C.5 in the supplementary material. The plots also allow a comparison of the prior (which is the first time point to the left) and the final posterior estimate (last time point on the right). The estimates stabilise nicely. We report about some limited validation of the parameter estimates in section D. Note that the variance of the random walk dynamics is estimated to be smaller than the one of the autoregressive model, as is also clear in Figure 2. On the other hand, the variance related to the piece-wise constant model is considerably smaller even if the variability seems to be smaller in Figure 2. This is due the fact that for most time points there are no discontinuities while when changes occur there may be large discontinuities.

In Figure 4 we use the estimated parameters, including the instantaneous reproduction numbers, to simulate the daily hospitalisation incidence and the daily number of positive tests. We propagate uncertainty and produce probabilistic estimates, which we compare with the actual data. These plots show that we are able to fit both data sources well. We note the weekly structure in the test data. These plots also show three weeks ahead forecasts. The superimposition of the actual data of these three weeks, which were not used in the analysis, show that predictions were good.
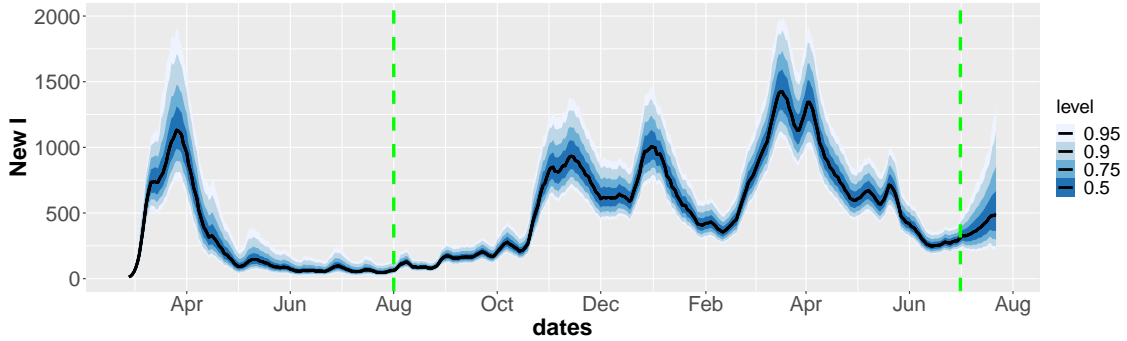
Figure 3: Predicted number of newly infected based on the autoregressive model in Figure 2. The estimates after the last dashed green line are predictions 3 weeks ahead.
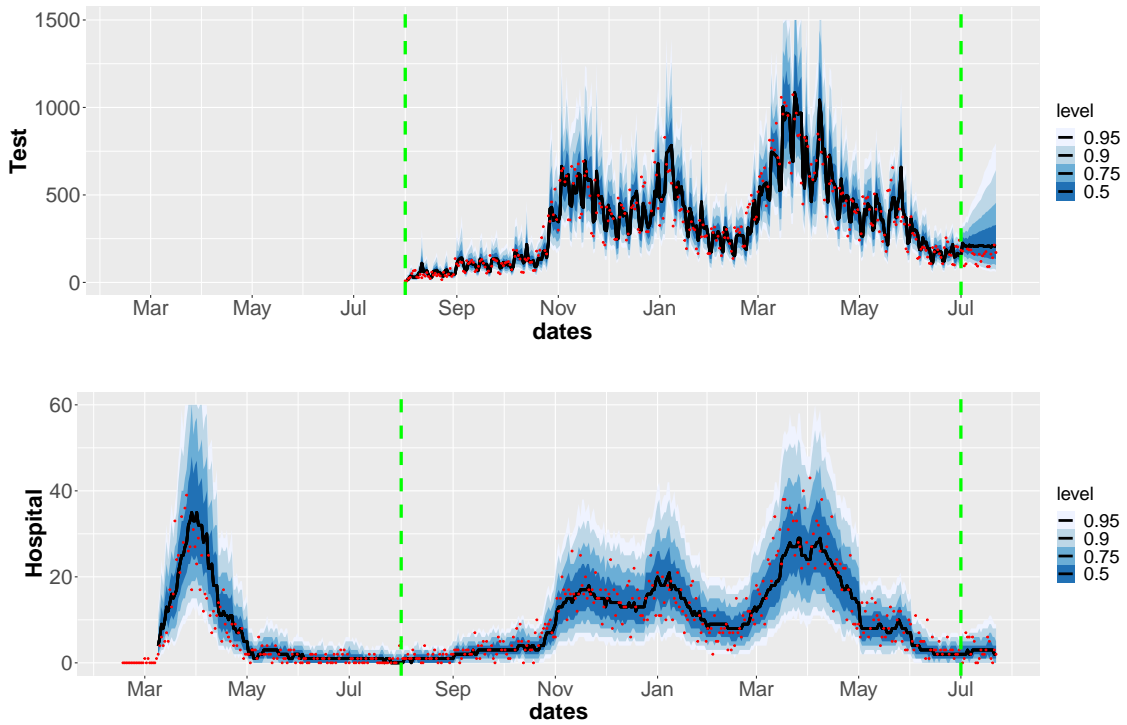

Figure 4: Predicted (blue shadows representing the quantiles of the posterior predictive distribution) and observed daily hospitalisation incidence and laboratory-confirmed positive tests incidence (red dots) based on the autoregressive model.

5.2 Sensitivity analysis

There are several parameters that are fixed in the current implementation:

- Hyper-parameters in the prior model for $\{R_t\}$, see Table 1.

- Parameters related to the observations: the age-dependent probabilities of being hospitalised and detected positive by testing, and the parameters describing the distribution for delay from infection to testing or hospitalisation.

- Several parameters inside the SEIR model.

- The number of particles used in the SMC runs.

We did not perform a systematic sensitivity analysis, but focus on the SMC design parameters. Figure C.1 shows posterior median estimates of the instantaneous reproduction numbers for several different prior settings for the autoregressive model. The figure indicates that the estimates of $\{R_t\}$ are not sensitive to the tested prior settings (credibility intervals also show similar results).

Next we studied the importance of the number of particles. Figure C.7 in the supplementary file compares results for the autoregressive model based on $B = 20\,000$ vs. $B = 2\,000$ particles. The difference in the estimated marginal likelihoods is small compared to the differences between models (about 5.0). This figure (and other similar ones, not included) shows that the results are quite stable and that $B = 2\,000$ might have been sufficient for estimation of $\{R_t\}$. However, if interest is also in the latent structure $\boldsymbol{s}_t$ or marginal likelihood values, more particles have shown to be necessary.

5.3 Comparison with EpiEstim

EpiEstim (Cori et al., 2013) is a popular method for estimation of the reproduction number R_t based on case incidence as well as imported cases. It does however not allow for the incorporation of multiple data, so that hospital data are not taken into account, nor mobility. In Figure 5 we compare estimates of R_t obtained by our SMC approach, though only using test data and this time from the start of the epidemic, with the results obtained by using the `EpiEstim` package in R (Cori et al., 2021, version 2.2.4). For `EpiEstim` we assumed a serial interval (the time between the onset of symptoms of a primary case and the onset of symptoms of secondary cases) with mean 7.5 days and a standard deviation of 3 days. We see a very good agreement between the two estimated curves in general. However, the 95% confidence bands are unrealistically narrower for `EpiEstim`. There are some differences between the estimates during March 2020 and during the summer 2020 and November 2020. There are also some small lag differences between the estimates obtained by our approach compared to the results obtained by `EpiEstim`. This is probably because the assumed models for the delay between infection and testing times are quite different. Furthermore, the effect of the importations of cases, which is not accounted for by `EpiEstim`, may be very relevant to explain the disagreement between the models in certain periods such as in January 2021 (Christmas) and in the end of March 2021.

5.4 Simulated data

Based on an estimate $\{R_t^*\}$ obtained by the AR model from the real data, we simulated $(\boldsymbol{s}_t, \boldsymbol{z}_t, \boldsymbol{y}_t)$ from the model for all t. Twenty independent data sets were generated and for each case the simulated set $\{y_t\}$ was used for estimating $\{R_t\}$. Figure 6 shows the results from one of these simulated data sets, demonstrating that the method is able to capture the main structure quite well, although in periods where the true R_t^* is considerably unstable (July/August 2020) estimates are worse. Figure C.6 in the supplementary material shows results based on all 20 data sets, demonstrating the strong information content in such type of data about the $\{R_t\}$ process.
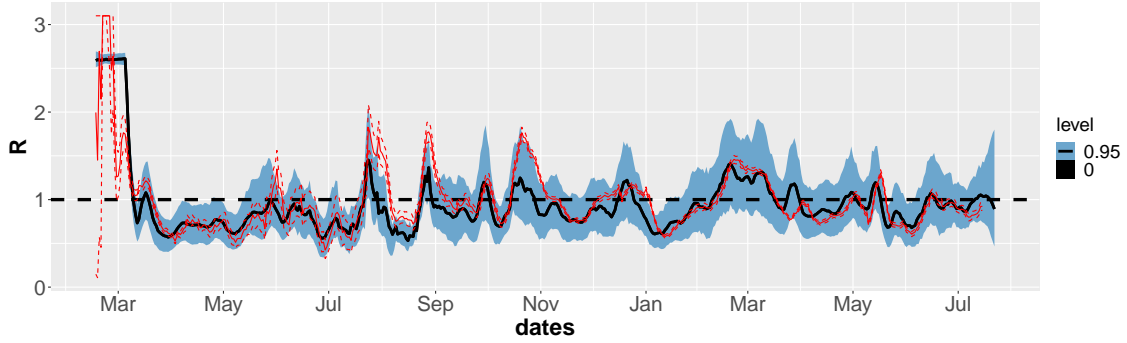
Figure 5: Estimates of R_t using the SEIR model combined with the autoregressive model for $\{R_t\}$ using SMC based only on test data (blue band, 95% confidence bands) and estimates obtained from `EpiEstim` (red lines, also 95% confidence bands).
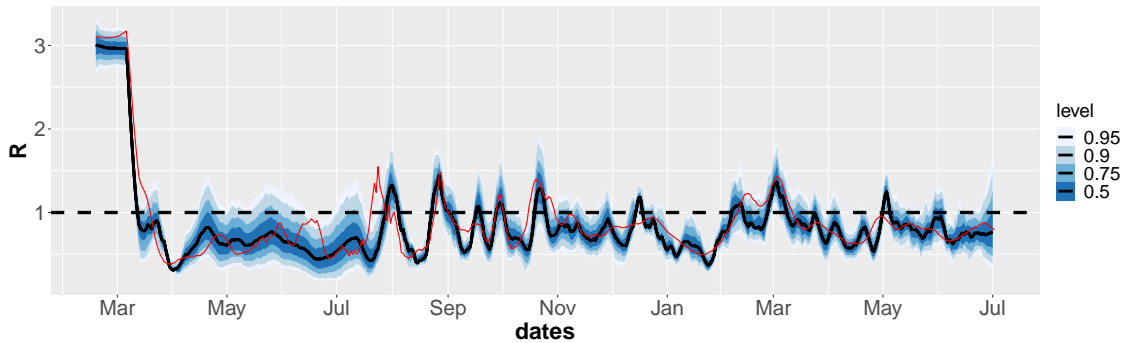

Figure 6: Simulation experiment. Estimates of $\frac{1}{7}\sum_{u=0}^{6} R_{t-u}$ using $B = 20\,000$ (blue) based on one simulated data set. Same setup as Figure 2, hospitalisation data from March 12 2020 to July 1 2021, test data from August 1 2020 to July 1 2021. The red curve corresponds to the assumed true $\{R_t^*\}$ process.

6   Discussion

A time-varying transmissibility allows to quantify the effects of interventions and changes in the behaviour of people, in real-time. This is of key importance to policy makers, as the interventions often have immense societal and health costs. Understanding, while an epidemic develops, whether the implemented interventions are sufficient or not, or if interventions could be lifted, is essential. The Norwegian government's strategy was to control the epidemic, and this was achieved by multiple national and local non-pharmaceutical interventions, which are reflected in the temporal variations of the reproduction number. To our knowledge, our approach is the first which allows to monitor a daily-varying reproduction number when using a complex compartmental model informed by multiple streams of data. The fact that our estimates of R_t react rapidly to changes in the test data, means that the situation is captured only with a delay given by the generation time of the disease under study and the time gap between transmission and testing. For Covid-19, this amounts to about a week, because of a generation time of about 5 days and a delay between transmission and test of about 2 days. Picking up an exponential growth ($R_t > 1$) before the epidemic grows out of control is essential for surveillance. The possibility of our method to validate the efficacy of contact tracing, to lead back R_t to below 1, or not, is also very important.

We have shown how daily reproduction numbers and the latent compartment-wise populations in an SEIR model can be put into a state space model, so that an SMC technique for inference can be used. Obtaining unbiased estimates of the marginal likelihoods also makes it possible to do parameter estimation within a particle MCMC framework, although more work is needed here to make this computationally efficient. Compared to a parallel effort using Approximate Bayesian Computation (ABC) (Engebretsen et al., 2021), the SMC approach is much faster and also easier to modify with respect to model changes, confirming the findings in Birrell et al. (2020). Our implementation is modular, does not depend on the specific epidemic model (here SEIR), so that alternatives can easily be tested.

So far only simple bootstrap filters have been applied. This can be improved by utilising more efficient proposals, alternative algorithms such as the resample-move algorithm (Gilks and Berzuini, 2001) which was used in Birrell et al. (2020) or the recent promising ideas of iterated auxiliary filters and twisted models (Guarniero et al., 2017; Heng et al., 2020). We expect these approaches to be very useful when we expand the SEIR model to have different reproductions numbers in each of the eleven Norwegian counties.

We compared three simple dynamics for R_t and found that the autoregressive model was slightly better than the others. More work needs to be done here to compare the models in terms of prediction. In our approach it is easy to predict the future hospitalisation incidence and the number of positive cases tested. Note however that such simple dynamic models for $\{R_t\}$ are mostly suitable for now-casting and for short term forecasting, because of the lack of stationary that interventions and feedbacks imply. There are then interesting questions on how to use our probabilistic prediction of the time varying transmission strength to propagate uncertainty, in the context of variable planned interventions.

Estimation of (static) parameters is a challenging task in SMC. Several parameters related to the SEIR model, as well as parameters related to the observation processes, were pre-estimated based on external data sources. In Engebretsen et al. (2021), we used a version of ABC. Parameters in the $\{R_t\}$ dynamic process were estimated online based on the procedure of sufficient statistics, a method that can lead to degeneracy. We have also tested out the particle Metropolis-Hastings algorithm by Andrieu et al. (2010). However convergence was slow and challenging, because of the computational cost of running the SMC algorithm even for a small number of particles. We validated our estimates in the supplementary material, and find that our online estimation procedure worked reasonably well for the given models and data. However, some experiments with the AR model when also including estimation of the parameter μ, did cause degeneracy problems as have been reported in in Andrieu et al. (2005). A more efficient SMC algorithm might be better in utilising the potential of such algorithms to estimate static parameters. However, our estimates of $\{R_t\}$ appear to be relatively robust with respect to changes in these parameters. Other parameters related to the SEIR model and the observation processes can be more important.

Communicating uncertainty of estimates and the effect of stochastic and uncertain time lags from data back to infections, is a major challenge. Our current strategy has been to report estimates of the reproduction numbers one week back in time as the most reliable estimates, and this needs to be studied further.

As pointed out by one of the our reviewers, it is quite easy to extend the modelling approach (and the algorithm) to include dummy variables describing interventions made by the government. In this case, one would need to estimate a time delay between the time point in which the interventions are decided and when they effect viral transmission. This delay might change in time and be intervention specific. In our model, interventions appear in the data after a delay, and are then reflected in a change of the reproduction number. It is possible to interpret changes in the estimated reproduction numbers in the light of the interventions set in place.

Finally, it is interesting to compare the estimates of the instantaneous reproduction number produced by this SMC model, with the estimates obtained by models which keep the reproduction number constant over longer time intervals, for example over four weeks. The SMC-based R_t is able to capture changes at shorter time scales significantly better, but possibly with larger uncertainty than estimates of reproduction numbers assumed to be constant over longer time periods, if the transmission has been stable during such periods. Comparing prediction power is a further aspects that can be examined. Results from our SMC model were used in the weekly reports of the Norwegian Institute of Public Health at least until April 2022, see www.fhi.no/en/publ/2020/weekly-reports-for-coronavirus-og-covid-19/.

Code and data

The data sets analysed in this paper come from the national emergency preparedness registry for Covid-19, owned by the Norwegian Institute of Public Health. The preparedness registry is temporary and comprises data from a variety of central health registries, national clinical registries and other national administrative registries. Further information on the registry, including access to data, is available at `www.fhi.no/en/id/infectious-diseases/coronavirus/emergency-preparedness-register-for-covid-19/`. See also `https://nordicmathcovid.cs.aalto.fi/data/`.

An R package called `smc.covid` is available at github.com/geirstorvik/smc.covid. It contain scripts for running the SMC algorithm on some data examples. Because data are sensitive, not all data sets are public. The number of individuals in hospital and the number of positive cases per day are available, but the number of imported cases and the mobility matrices are confidential. For the two latter data sets we have therefore provided some simulated synthetic data sets which resemble the true ones.

Acknowledgement

References

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

Andrieu, C., A. Doucet, and V. B. Tadic (2005). On-line parameter estimation in general state-

space models. In *Proceedings of the 44th IEEE Conference on Decision and Control*, pp. 332–337. IEEE.

Andrieu, C., G. O. Roberts, et al. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *The Annals of Statistics 37*(2), 697–725.

Birrell, P. J., L. Wernisch, B. D. Tom, L. Held, G. O. Roberts, R. G. Pebody, D. De Angelis, et al. (2020). Efficient real-time monitoring of an emerging influenza pandemic: How feasible? *Annals of Applied Statistics 14*(1), 74–93.

Blackwood, J. C., D. A. Cummings, H. Broutin, S. Iamsirithaworn, and P. Rohani (2013). Deciphering the impacts of vaccination and immunity on pertussis epidemiology in thailand. *Proceedings of the National Academy of Sciences 110*(23), 9595–9600.

Cauchemez, S., P.-Y. Boëlle, G. Thomas, and A.-J. Valleron (2006). Estimating in real time the efficacy of measures to control emerging communicable diseases. *American journal of epidemiology 164*(6), 591–597.

Chopin, N. and O. Papaspiliopoulos (2020). *An introduction to sequential Monte Carlo.* Springer Series in Statistics. Springer.

Cori, A., N. M. Ferguson, C. Fraser, and S. Cauchemez (2013). A new framework and software to estimate time-varying reproduction numbers during epidemics. *American journal of epidemiology 178*(9), 1505–1512.

Cori, A., Z. Kamvar, J. Stockwin, T. Jombart, E. Dahlqwist, R. FitzJohn, and R. Thompson (2021). EpiEstim v2.2-3: A tool to estimate time varying instantaneous reproduction number during epidemics. https://github.com/mrc-ide/EpiEstim.

De Angelis, D., A. M. Presanis, P. J. Birrell, G. S. Tomba, and T. House (2015). Four key challenges in infectious disease modelling using data from multiple sources. *Epidemics 10*, 83–87.

Dunbar, M. B.-N. and L. Held (2021). Assessing the effect of school closures on the spread of covid-19 in zurich. Special topic meeting on Covid-19 transmission.

Engebretsen, S., K. Engø-Monsen, M. A. Aleem, E. S. Gurley, A. Frigessi, and B. F. de Blasio (2020). Time-aggregated mobile phone mobility data are sufficient for modelling influenza spread: the case of Bangladesh. *Journal of the Royal Society Interface 17*(167), 1–14.

Engebretsen, S., K. Engø-Monsen, A. Frigessi, and B. F. de Blasio (2019). A theoretical single-parameter model for urbanisation to study infectious disease spread and interventions. *PLoS computational biology 15*(3), e1006879.

Engebretsen, S., A. N. Osnes, G. Rø, and R. White (2020). *spread: Infectious Disease Spread Models.* Norwegian Institute of Public Health. R package version 2020.9.16.

Engebretsen, S., G. Rø, A. D.-L. Palomares, K. Engø-Monsen, A. B. Kristoffersen, F. Di Ruscio, A. Frigessi, and B. F. de Blasio (2021). Spatial modelling of the early-phase of the COVID-19 epidemic in Norway. https://www.medrxiv.org/content/10.1101/2021.10.25.21265166v1. medRxiv, Cold Spring Harbor Laboratory Press.

Fearnhead, P. (2002). Markov chain Monte Carlo, sufficient statistics, and particle filters. *Journal of Computational and Graphical Statistics 11*(4), 848–862.

Ferretti, L., C. Wymant, M. Kendall, L. Zhao, A. Nurtay, L. Abeler-Dörner, M. Parker, D. Bonsall, and C. Fraser (2020). Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science 368*(6491), eabb6936.

Flaxman, S., S. Mishra, A. Gandy, H. J. T. Unwin, T. A. Mellan, H. Coupland, C. Whittaker, H. Zhu, T. Berah, J. W. Eaton, et al. (2020). Estimating the effects of non-pharmaceutical interventions on covid-19 in europe. *Nature 584*(7820), 257–261.

Gibson, G. J. and E. Renshaw (1998). Estimating parameters in stochastic compartmental models using Markov chain methods. *Mathematical Medicine and Biology: A Journal of the IMA 15*(1), 19–40.

Gilks, W. R. and C. Berzuini (2001). Following a moving target—Monte Carlo inference for dynamic Bayesian models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63*(1), 127–146.

Guarniero, P., A. M. Johansen, and A. Lee (2017). The iterated auxiliary particle filter. *Journal of the American Statistical Association 112*(520), 1636–1647.

Heng, J., A. N. Bishop, G. Deligiannidis, A. Doucet, et al. (2020). Controlled sequential Monte Carlo. *Annals of Statistics 48*(5), 2904–2929.

Kantas, N., A. Doucet, S. S. Singh, J. Maciejowski, N. Chopin, et al. (2015). On particle methods for parameter estimation in state-space models. *Statistical science 30*(3), 328–351.

Keeling, M. J. and P. Rohani (2011). *Modeling infectious diseases in humans and animals*. Princeton university press.

King, A. A., D. Nguyen, and E. L. Ionides (2016). Statistical Inference for Partially Observed Markov Processes via the R Package pomp. *Journal of Statistical Software 69*(12), 1–43.

Liu, J. S. and R. Chen (1998). Sequential Monte Carlo methods for dynamic systems. *Journal of the American statistical association 93*(443), 1032–1044.

Mishra, S., J. Scott, H. Zhu, N. M. Ferguson, S. Bhatt, S. Flaxman, and A. Gandy (2020). A covid-19 model for local authorities of the united kingdom. *Medrxiv*.

NIPH (2022). Weekly reports for coronavirus and COVID-19. https://www.fhi.no/en/publ/2020/weekly-reports-for-coronavirus-og-covid-19/. Norwegian Institute of Public Health.

O'Neill, P. D., D. J. Balding, N. G. Becker, M. Eerola, and D. Mollison (2000). Analyses of infectious disease data from household outbreaks by Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 49*(4), 517–542.

O'Neill, P. D. and G. O. Roberts (1999). Bayesian inference for partially observed stochastic epidemics. *Journal of the Royal Statistical Society: Series A (Statistics in Society) 162*(1), 121–129.

Parag, K. V., R. N. Thompson, and C. A. Donnelly (2021). Are epidemic growth rates more informative than reproduction numbers? *medRxiv*.

Pellis, L., P. J. Birrell, J. Blake, C. E. Overton, F. Scarabel, H. B. Stage, and E. Brooks-Pollock (2022). Estimation of reproduction numbers in real time: conceptual and statistical challenges. *Journal of the Royal Statistical Society Series A.(to appear)*.

RSS (2021). Special topic meeting on Covid-19 transmission. https://rss.org.uk/news-publication/publications/journals/special-topic-meeting-on-r/. Royal Statistical Society, accessed: 2022-04-05.

Stocks, T., T. Britton, and M. Höhle (2020). Model selection and parameter estimation for dynamic epidemic models via iterated filtering: application to rotavirus in germany. *Biostatistics 21*(3), 400–416.

Storvik, G. (2002). Particle filters for state-space models with the presence of unknown static parameters. *IEEE Transactions on signal Processing 50*(2), 281–289.

Teh, Y. W., A. Bhoopchand, P. Diggle, B. Elesedy, B. He, M. Hutchinson, U. Paquet, J. Read, N. Tomasev, and S. Zaidi (2021). Efficient Bayesian Inference of Instantaneous Reproduction Numbers at Fine Spatial Scales, with an Application to Mapping and Nowcasting the Covid-19 Epidemic in British Local Authorities. *URL https://rss. org. uk/RSS/media/File-library/News/2021/WhyeBhoopchand. pdfhttps://localcovid. info/2*.

Viboud, C., K. Sun, R. Gaffey, M. Ajelli, L. Fumanelli, S. Merler, Q. Zhang, G. Chowell, L. Simonsen, A. Vespignani, and the RAPIDD Ebola Forecasting Challenge group (2018). The rapidd ebola forecasting challenge: Synthesis and lessons learnt. *Epidemics 22*, 13–21.

Whittaker, R., A. B. Kristofferson, E. Seppälä, B. V. Salamanca, L. Veneti, M. L. Storm, H. Bøås, N. Aasand, U. Naseer, K. Bragstad, et al. (2021). Trajectories of hospitalisation for patients infected with SARS-CoV-2 variant B. 1.1. 7 in Norway, December 2020–April 2021. *Journal of Infection 83*(4), e14–e17.

Supplementary material: A sequential Monte Carlo approach to estimate a time varying reproduction number in infectious disease models: the Covid-19 case.

Geir Storvik[1]   Alfonso Diz-Lois Palomares[2]   Solveig Engebretsen[3]
Gunnar Øyvind Isaksson Rø[2]   Kenth Engø-Monsen[4]
Anja Bråthen Kristoffersen [2]   Birgitte Freiesleben de Blasio[2,5]
Arnoldo Frigessi[5,6]

May 3, 2022

A Derivations of the dynamic models for $\{R_t\}$

Here we define $L_t = \log(R_t)$ for simplicity of notation and neglect the first time points (before March 7 2020) for which R_t is constant. Note that for each model, the sufficient statistics involved are all easily updated with increasing t.

A.1 The random walk model

Assume a prior $p(\tau) = \text{Gamma}(\alpha_\tau, \beta_\tau)$ with $\tau = 1/\sigma_R^2$. Given $\{R_t\}$, the posterior for (ϕ_c, τ) becomes

$$p(\tau|\boldsymbol{R}_{1:t}) \propto \tau^{\alpha_\tau - 1} \exp(-\beta_\tau \tau) \prod_{s=1}^{t} \tau^{1/2} \exp(-\tfrac{1}{2}\tau(L_s - L_{s-1})^2)$$

$$\propto \tau^{\alpha_\tau + 0.5t} \exp(-[\beta_\tau + 0.5 \sum_{s=1}^{t}(L_s - L_{s-1})^2]\tau)$$

$$\propto \text{Gamma}(\tau; \alpha_\tau + 0.5t, \beta_\tau + 0.5 \sum_{s=1}^{t}(L_s - L_{s-1})^2).$$

Simulation of L_{t+1} can then be performed by first simulating τ from the Gamma distribution and thereafter R_{t+1} from the model.

A.2   The piecewise constant model

Assume a prior $p(\phi_c, \tau) = \text{Beta}(a_\phi, b_\phi)\text{Gamma}(\alpha_\tau, \beta_\tau)$ with $\tau = 1/\sigma_R^2$. Given $\{R_t\}$ and change-point indicators C_t ($= 1$ if there is a changepoint at time t), the posterior for (ϕ_c, τ) becomes

$$
\begin{aligned}
p(\phi_c, \tau | \boldsymbol{R}_{1:t}) \propto & \phi_c^{a_\phi - 1}(1 - \phi_c)^{b_\phi - 1}\tau^{\alpha_\tau - 1}\exp(-\beta_\tau \tau) \times \\
& \prod_{s=1}^{t}[(1 - \phi_c)^{1 - C_s}\phi_c^{C_s}\tau^{C_s/2}\exp(-\tfrac{1}{2}\tau C_s(L_s - L_{s-1})^2)] \\
\propto & \phi_c^{a_\phi + n_t^c - 1}(1 - \phi_c)^{b_\phi + t - n_t^c - 1}(1 - \phi_c)^{n_t^c} \times \\
& \tau^{\alpha_\tau + 0.5 n_t^c}\exp(-[\beta_\tau + 0.5\sum_{s=1}^{t}C_s(L_s - L_{s-1})^2]\tau) \\
\propto & \text{Beta}(\phi_c; a_\phi + n_t^c, b_\phi + t - n_t^c) \times \\
& \text{Gamma}(\tau; \alpha_\tau + 0.5 n_t^c, \beta_\tau + 0.5\sum_{t=1}^{T}C_s(L_s - L_{s-1})^2)
\end{aligned}
$$

where n_t^c is the number of changepoints up to timepoint t. Simulation of R_{t+1} can then be performed by first simulating (ϕ_c, τ) from the Beta-Gamma distribution and thereafter R_{t+1} from the model.

A.3   The autoregressive model

We consider here only the case where $\mu = 0$. Assume

$$
p(a, \tau) = N(a_0, [\kappa_0 \tau]^{-1}) \times \text{Gamma}(\alpha_\tau, \beta_\tau)
$$

Note first that

$$
\begin{aligned}
g_t(a) \equiv & \kappa_0(a - a_0)^2 + \sum_{s=1}^{t}(L_s - aL_{s-1})^2 \\
= & \kappa_0 a^2 - 2\kappa_0 a_0 a + \kappa_0 a_0^2 + \sum_{s=1}^{t}L_s^2 - 2a\sum_{s=1}^{t}L_s L_{s-1} + a^2\sum_{s=1}^{t}L_{s-1}^2 \\
= & (\kappa_0 + \sum_{s=1}^{t}L_{s-1}^2)a^2 - 2[\kappa_0 a_0 + \sum_{s=1}^{t}L_s L_{s-1}]a + \kappa_0 a_0^2 + \sum_{s=1}^{t}L_s^2 \\
= & (\kappa_0 + \sum_{s=1}^{t}L_{s-1}^2)(a - \hat{a}_t)^2 - (\kappa_0 + \sum_{s=1}^{t}L_{s-1}^2)\hat{a}_t^2 + \kappa_0 a_0^2 + \sum_{s=1}^{t}L_s^2
\end{aligned}
$$

with

$$
\hat{a}_t = \frac{\kappa_0 a_0 + \sum_{s=1}^{t}L_s L_{s-1}}{\kappa_0 + \sum_{s=1}^{t}L_{s-1}^2}.
$$

This gives

$$
\begin{aligned}
p(a | \tau, R_{1:t}) \propto & \exp\left(-0.5\tau g_t(a)\right) \\
\propto & N\left(\hat{a}_t, \tau^{-1}[\kappa_0 + \sum_{s=1}^{t}L_{s-1}^2]^{-1}\right).
\end{aligned}
$$

Further, for any value of a, we have

$$
p(\tau | R_{1:t}) \propto \frac{p(\tau)p(R_{1:t} | \tau, a)}{p(a | \tau, R_{1:t})}
$$

2

Prob	0.0378	0.0230	0.0209	0.0180	0.0146	0.0185
From day	2020-02-17	2020-05-01	2020-06-01	2020-07-01	2020-08-01	2020-09-01
Prob	0.0183	0.0208	0.0232	0.0221	0.0230	0.0224
From day	2020-10-01	2020-11-01	2020-12-01	2021-01-01	2021-02-01	2021-03-01
Prob	0.0231	0.0141				
From day	2021-04-01	2021-05-01				

Table 1: Time-changing probabilities for hospitalisation, given infected. The dates correspond to starting day for the probability provided.

and using $a = \hat{a}_t(a)$, we obtain

$$p(\tau|R_{1:t}) \propto \frac{\tau^{\alpha_\tau - 1} \tau^{0.5t} \exp(-\beta_\tau \tau) \exp(-0.5\tau g_t(\hat{a}))}{\tau^{0.5}}$$
$$\propto \text{Gamma}(\alpha_\tau + 0.5t, \beta_\tau + 0.5g_t(\hat{a}_t)).$$

B   Parameter settings

The full model includes several parameter specifications that are fixed throughout all runs. Some sensitivity analysis is provided in Section C.

- R_t was set to a constant R_0 from February 17 2020 to March 8 2020. We estimate R_0 and used a $N(1.192, 0.1)$ prior for $\log R_0$ (giving an expectation of R_0 equal to 3.3, which is compatible with the literature).

- The amplification factor for imported cases is set to 2.8 from (Engebretsen et al., 2021) .

- In order to take into account regional variability, the number of hospitalised individuals is assumed to follow a beta-binomial distribution with expectations equal to the probabilities given in Table 1 and with a shape parameter equal to 8. The time-dynamic changes in the probabilities are based on (Engebretsen et al., 2021).

- The distribution for the delay from infected to hospitalised is assumed to follow a negative binomial distribution with parameters $(8.87, 3.40)$ until August 1 2020 and thereafter with parameters $(7.60, 3.34)$, as estimated from registry data.

- The distribution of the time between infection and testing positive is assumed to follow a discrete distribution with probabilities $0.05, 0.080, 0.16, 0.4, 0.3, 0.01$ for delays of 1-6 days.

- In the model (5) for the detection probability, we used $\pi_0 = -1.017, \pi_1 = 0.00013$. The number of detected individuals was assumed to follow a beta-binomial distribution with expectations given by (5) and a shape parameter equal to 8.

C   Additional results

Figure C.1 explores the sensitivity of results with respect to some of the parameters given in Section B. In this figure (where only posterior medians are shown), the black line corresponds to weekly averaged fixed lag estimates of $\{R_t\}$ based on the autoregressive model (the middle panel in Figure 2). For the red curve, we changed the prior for $\tau = 1/\sigma^2$ to a Gamma$(1.2, 0.14)$ distribution. For the blue curve, we changed the values of (π_0, π_1) to (-1.175,0.000074) (estimates obtained from another model on the same data). The green curve corresponds to a change in the shape parameters in the negative binomial distribution for hospitalisation and test data from 8 to 3. In all cases the results are stable, which was also the case when we compared 50% credibility intervals (not shown).

Figure C.2 shows the fixed-24-days-lag smoothed estimates of daily R_t values for the three models. This plot is not averaged over each preceding week, as is done in Figure 2 in the main text.

	Repetitions					Mean	RW	AR	CP
RW	-3031.56	-3029.18	-3033.91	-3030.74	-3038.64	-3032.81		0.0000	0.0096
AR	-3008.85	-3008.06	-3006.69	-3010.22	-3014.15	-3009.60			0.0002
CP	-3026.68	-3029.23	-3024.41	-3024.04	-3018.43	-3024.56			

Table 2: On the left, estimates of marginal log-likelihoods based on $B = 5\,000$ particles using five repetitions of the SMC algorithm. On the right, p-values based on a t-test for testing the hypothesis on equality for the marginal log-likelihoods between the models.
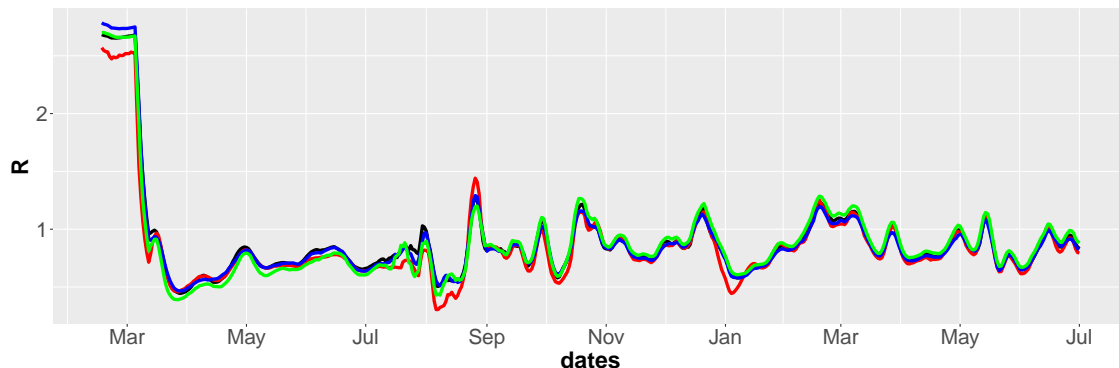
Figure C.1: Fixed lag estimates of R_t for different prior settings for the autoregressive model. Black corresponds to the settings described in section B and priors given in Table 1, the red curve is obtained by changing (π_0, π_1) to (-1.175,0.000074), the green curve is obtained by changing the prior for τ to Gamma$(1.2, 0.14)$ and the blue curve corresponds to changing the α parameters in the Negative-binomial distributions for hospital and test data to 4.

Figures C.3-C.5 show (filtered) parameter estimates of the dynamics of $\{R_t\}$ for the three models considered. All parameter estimates are based on the use of sufficient statistics (Fearnhead, 2002; Storvik, 2002) in the SMC algorithm.

Table 2 shows that although there is some Monte Carlo variability in the estimation of the marginal likelihoods, the variability is small enough to capture differences between the different models considered.

Figure C.6 summarises results from the simulation experiments reported in Section 5.4. For each of the 20 simulation experiments (with the same underlying true $\{R_t\}$ process), posterior medians were obtained similar as to the runs on the real data. The figure shows the variability in these medians over different data sets, demonstrating the strong information content in such data about the $\{R_t\}$ process.

Figure C.7 gives a comparison of one run based on $B = 20\,000$ particles and one run with $B = 2\,000$ particles, indicating that the number of particles used are sufficient for reliable estimates of the $\{R_\}$ process.

D   Estimation of hyper-parameters

As described in Kantas et al. (2015), there are many different approaches to estimate static hyper-parameters, both online and offline. In the paper we focused on online methods for estimation of parameters in the $\{R_t\}$ process, based on the use of sufficient statistics (Fearnhead, 2002; Storvik, 2002). As pointed out by one of the referees, this method can suffer from degeneracy and converge towards different limit distributions in different runs. Such degeneracy problems can in particular occur for long time series. We validate the estimates obtained in two ways. We concentrate on the AR process, both due to the fact that this is our best model and that it turned out to be the most problematic one, with respect to parameter estimation.
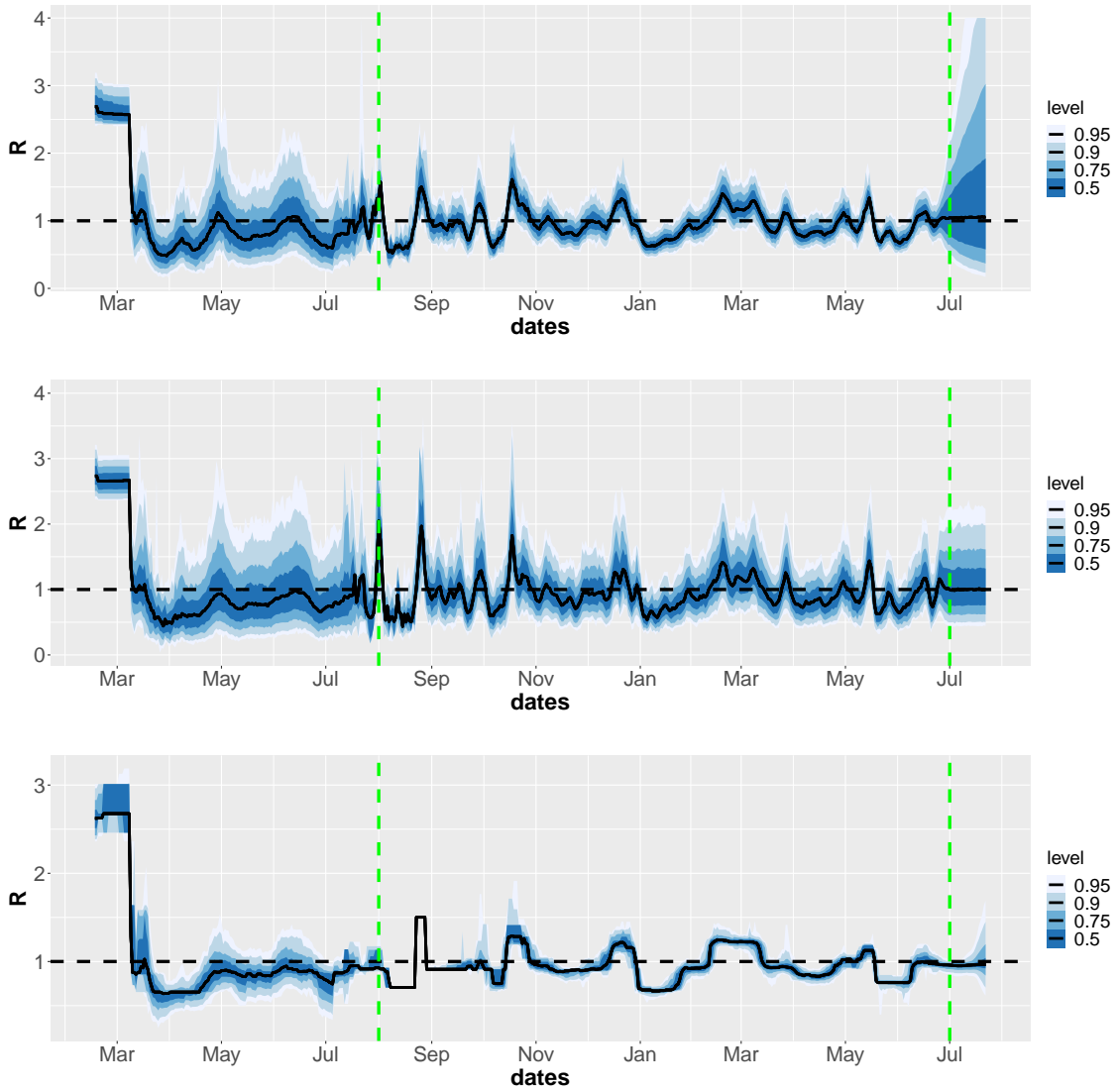
4

Figure C.2: Estimates of R_t based on the random walk (upper), autoregressive (middle) and piece-wise constant models. Corresponding weekly averaged estimates are given in Figure 2.
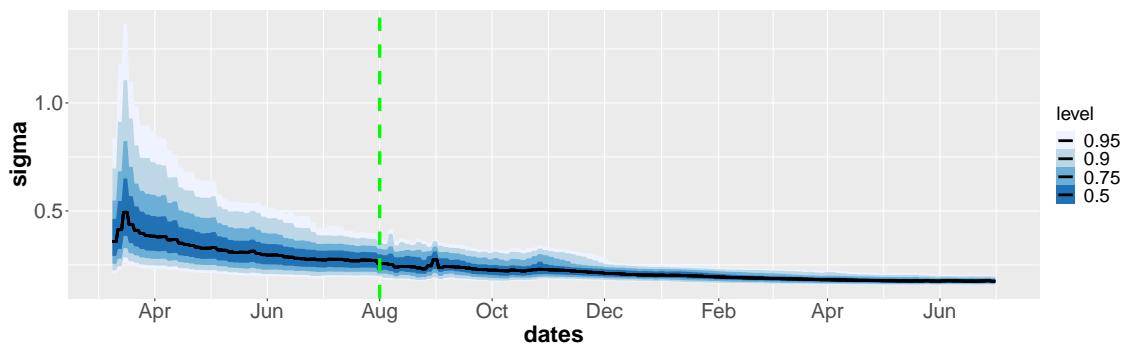
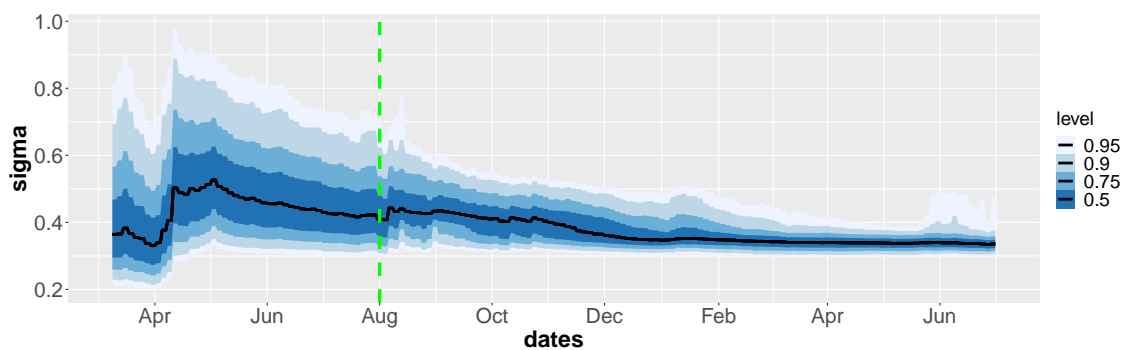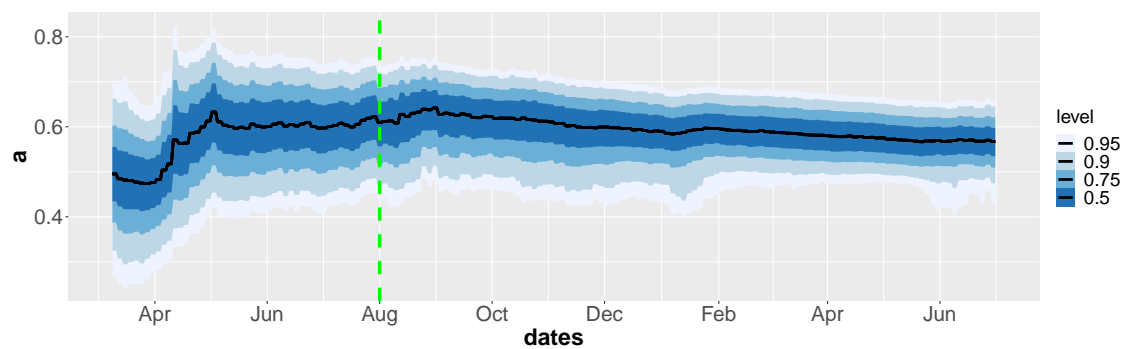Figure C.3: Filtered estimates of σ for the random walk model.

Figure C.4: Filtered estimates of a (top) and σ (bottom) for the autoregressive model.
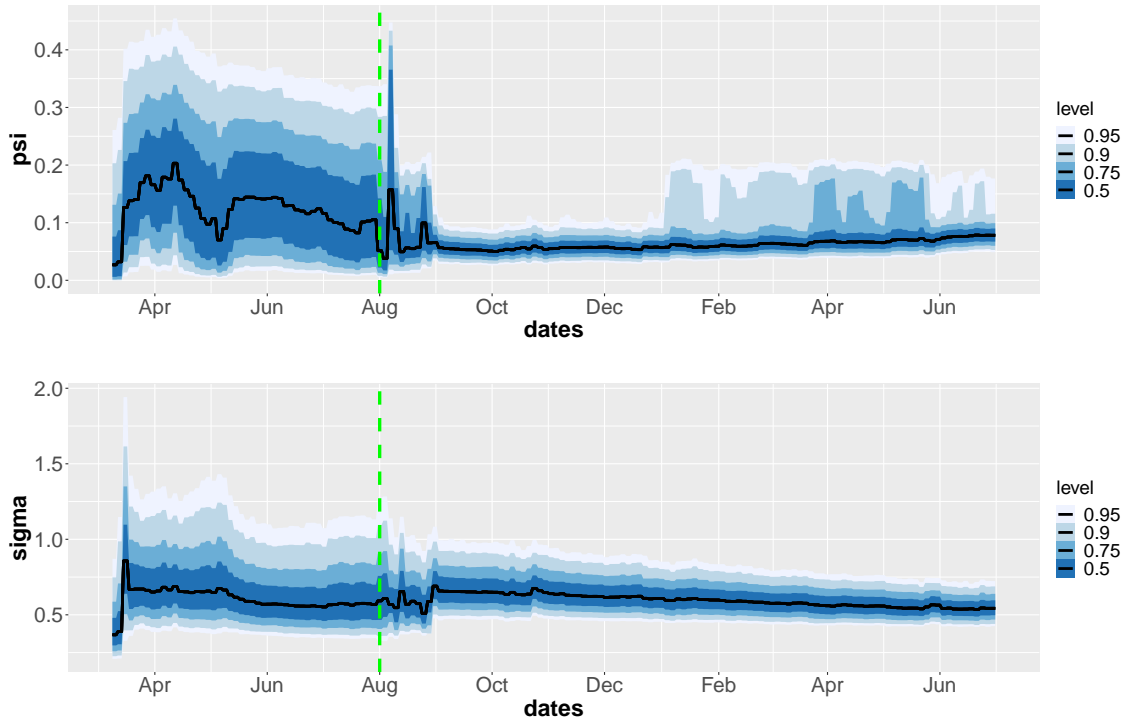
Figure C.5: Filtered estimates of ϕ_c (top) and σ_R (bottom) for the piecewise constant model.
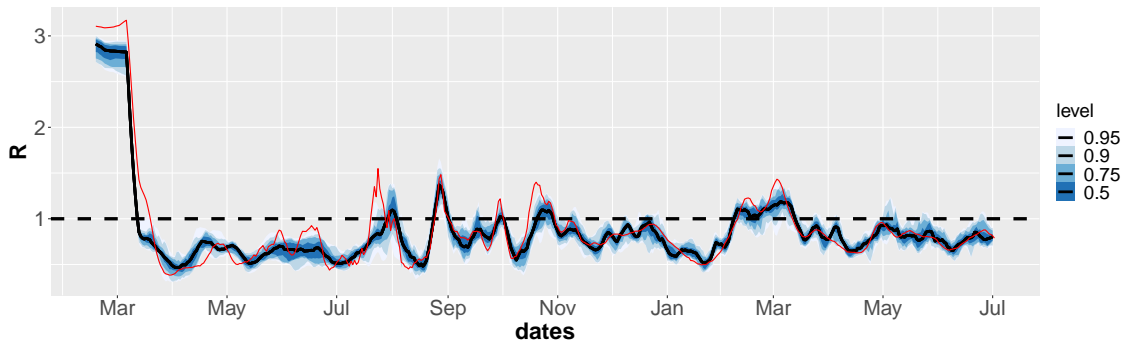

Figure C.6: Results from 20 independent simulation experiments. All 20 datasets simulated based on the same true $\{R_t\}$ process (red line). The blue bands show credibility intervals of the 20 posterior medians.
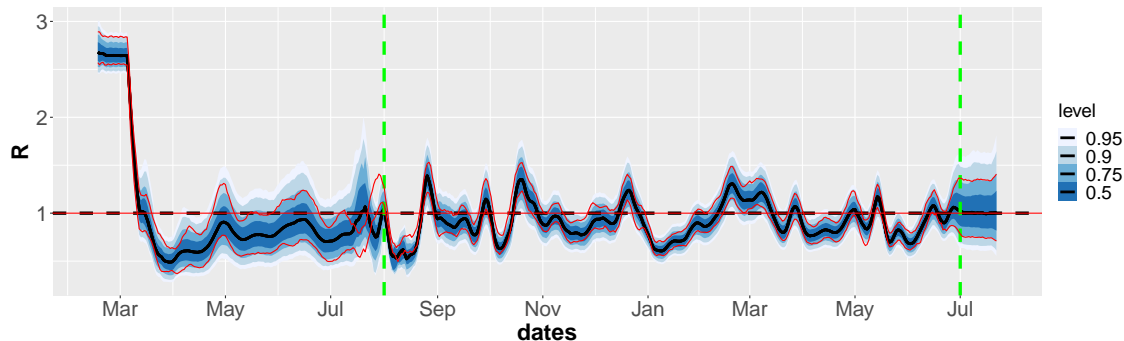
Figure C.7: Estimates of $\frac{1}{7}\sum_{u=0}^{6} R_{t-u}$ using $B = 20\,000$ (blue) and $B = 2\,000$ (red) particles based on the autoregressive model. Same setup as for Figure 2.

Figure D.1 shows the parameter estimates obtained from two independent runs. The agreement of these runs indicates that the estimation procedure performs quite well in our case. We do however emphasise that even though the problems with degeneracy did not seem to be present for the model considered here, if we considered an AR process where also μ is estimated, indeed different runs of the SMC algorithm resulted in somewhat different parameter estimates. The problems in that case occurred on the day where test data were introduced.

An important feature of SMC methods is that unbiased estimates of marginal likelihoods are available, making the pseudo-marginal approach by Andrieu et al. (2009) possible. Given a current parameter $\boldsymbol{\theta}$ with corresponding estimate of the marginal likelihood $\hat{p}(\boldsymbol{y}_{1:T}; \boldsymbol{\theta})$, a proposal $\tilde{\boldsymbol{\theta}} \sim q(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta})$ is generated and an estimate $\hat{p}(\boldsymbol{y}_{1:T}; \tilde{\boldsymbol{\theta}})$ is obtained by a new run of the SMC algorithm. The new proposal will then be accepted with probability

$$\min\{1, \frac{p(\tilde{\boldsymbol{\theta}})\hat{p}(\boldsymbol{y}_{1:T}; \tilde{\boldsymbol{\theta}})q(\boldsymbol{\theta}|\tilde{\boldsymbol{\theta}})}{p(\boldsymbol{\theta})\hat{p}(\boldsymbol{y}_{1:T}; \boldsymbol{\theta})q(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta})}\},$$

corresponding to the Particle Metropolis-Hastings Algorithm in Andrieu et al. (2010). Practical use of this algorithm requires tuning of the proposal distribution $q(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta})$ and of the number of particles used in each SMC run. Concerning the latter, the number of particles in each SMC run needs to be reduced dramatically compared to a single run of the SMC algorithm since these runs have to be repeated many times. Although Andrieu et al. (2010) showed that the number of particles should increase with time, we have only considered a fixed number of particles in our testing. Even with many trials on the tuning parameters, we struggled to obtain convergence within reasonable time (running the algorithm for several days). However, we stored all the proposed parameter settings together with their likelihood values. We then fitted a two-dimensional GAM model to the log-likelihood values using a and σ as covariates. More than 500 samples of (a, σ) were obtained through multiple runs of the PMCMC algorithm with different starting points (running in total for several days). Figure D.2 shows the fitted GAM curve with the red circles corresponding to the samples obtained from the SMC algorithm. The SMC samples seem to be within the higher likelihood region of the fitted curve. Note that due to lack of converge, the PMCMC algorithm has probably not been able to explore the full posterior properly. Further, the "observed" values displayed as small dots in the plot are *all* proposed values, not only the accepted ones, so these points should not be used to reflect the whole distribution.

References

Andrieu, C., A. Doucet, and R. Holenstein (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72*(3), 269–342.
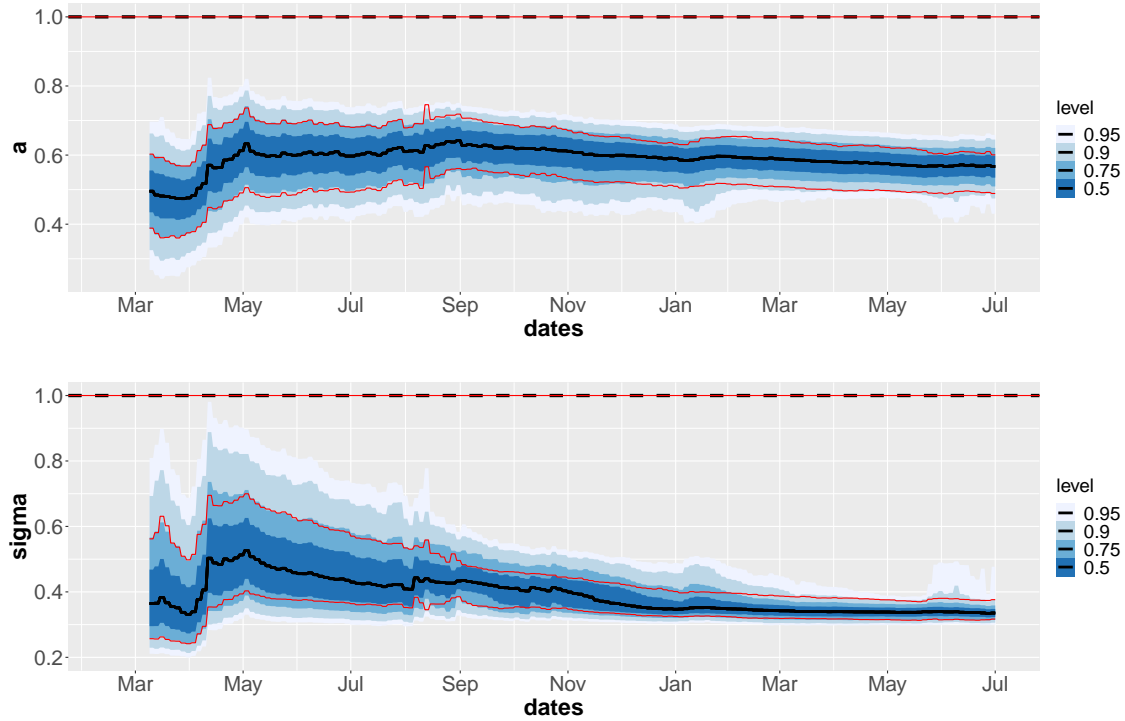
Figure D.1: Estimates of a (top) and σ (bottom) for the AR model. The blue bands show different confidence bands while the red curves show 95% confidence bands from an independent run of the SMC algorithm. Settings are the same as for the results in Figure 2.
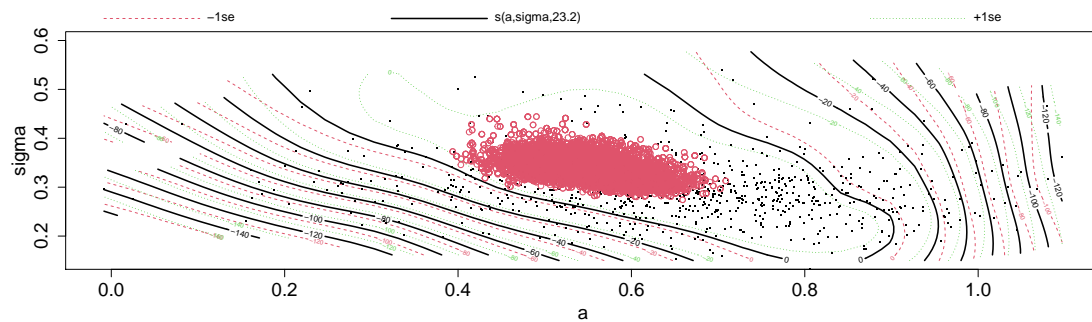

Figure D.2: Fitted two-dimensional GAM curve estimating the log-likelihood function based on values obtained through a run of the Particle Metropolis-Hastings algorithm. $B = 250$ particles were used for each run of the SMC algorithm, the proposal distributions were $N(a_{cur}, 1)$ for a and $N(\log \sigma_{cur}, 0.125)$ for σ on log-scale. Merged results from several runs of PMCMC with different starting points.

Andrieu, C., G. O. Roberts, et al. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *The Annals of Statistics 37*(2), 697–725.

Engebretsen, S., G. Rø, A. D.-L. Palomares, K. Engø-Monsen, A. B. Kristoffersen, F. Di Ruscio, A. Frigessi, and B. F. de Blasio (2021). Spatial modelling of the early-phase of the COVID-19 epidemic in Norway. https://www.medrxiv.org/content/10.1101/2021.10.25.21265166v1. medRxiv, Cold Spring Harbor Laboratory Press.

Fearnhead, P. (2002). Markov chain Monte Carlo, sufficient statistics, and particle filters. *Journal of Computational and Graphical Statistics 11*(4), 848–862.

Kantas, N., A. Doucet, S. S. Singh, J. Maciejowski, N. Chopin, et al. (2015). On particle methods for parameter estimation in state-space models. *Statistical science 30*(3), 328–351.

Storvik, G. (2002). Particle filters for state-space models with the presence of unknown static parameters. *IEEE Transactions on signal Processing 50*(2), 281–289.