PAPER

Bayesian particle filter algorithm for learning epidemic dynamics

To cite this article: D Calvetti *et al* 2021 *Inverse Problems* **37** 115008

IOP Publishing Inverse Problems
Inverse Problems **37** (2021) 115008 (30pp) https://doi.org/10.1088/1361-6420/ac2cdc

Bayesian particle filter algorithm for learning epidemic dynamics

D Calvetti[1,2,*], A Hoover[3], J Rose[2] and E Somersalo[1,2]

[1] Department of Mathematics, Applied Mathematics, and Statistics, Case Western Reserve University, Cleveland, Ohio, United States of America
[2] Center for Community Health Integration, Case Western Reserve University, Cleveland, Ohio, United States of America
[3] Department of Mathematics, University of Akron, Akron, Ohio, United States of America

E-mail: dxc57@case.edu

Received 8 February 2021, revised 8 August 2021
Accepted for publication 5 October 2021
Published 22 October 2021

Abstract

In this article, we consider a dynamic model for the spread of epidemics, in particular of COVID-19, and the inverse problem of estimating sequentially the time evolution of the unknown state and the model parameters based on noisy observations of the new daily infections. A characteristic of COVID-19 is the significant proportion of secondary infections though contacts with asymptomatic or oligosymptomatic infectious individuals. Since most of these individuals are not accounted for in the number of new daily infections, the size of this cohort can be inferred only indirectly through the underlying model. The evolution model used to propagate the current state from one data instance to the next is a suitably modified SEIR compartment model, providing the expected value for the new daily infection count that is modeled as a Poisson distributed random variable. The estimation of the state and the model parameters is based on a Bayesian particle filtering algorithm. The sequential Bayesian framework naturally provides a quantification of the uncertainty in the estimates of the model parameters, basic reproduction number, and size of the cohorts. Of particular interest is the fact that the algorithm makes it possible to estimate the size of the asymptomatic cohort, a key component for understanding the COVID-19 dynamics, and for planning mitigation measures. Alternative versions of the classical basic reproduction number for estimating the speed of the propagation of the disease are also proposed. The viability of the algorithm is demonstrated through a set of computed examples with both simulated realistic data and actual real data from selected US counties. The numerical tests show that the algorithm

*Author to whom any correspondence should be addressed.

reproduces a ratio of asymptomatic vs symptomatic cohort sizes remarkably close to what is currently suggested by the Center for Disease Control.

Keywords: COVID-19, basic reproduction number, state estimation, parameter estimation, sequential Monte Carlo

(Some figures may appear in colour only in the online journal)

1. Introduction

In a little over four months since its emergence in Wuhan, China, at the end of 2019, the novel coronavirus SARS-CoV-2, the cause of COVID-19, spread worldwide, evolving into a global pandemic in spite of the measures taken to control and contain the contagion. SARS-CoV-2 is itself part of the family of coronaviruses that were responsible for the SARS and MERS epidemics in 2003 and 2009, respectively, though there are significant differences between these epidemics and the current COVID-19 pandemic. One major difference is the asymptomatic/presymptomatic transmission of COVID-19, which has turned out to be a major hurdle for containing the spread. Consequently, existing mathematical models for understanding the dynamics of the epidemics needed to be updated to better account for the specific features of the disease.

Since the dawn of mathematical epidemiology [1], SIR and SEIR compartment models have played a central role in understanding the spread of infectious diseases, and the estimation of the related model parameters from available data remains a central issue. The spread, speed, and severity of the impact of the COVID-19 pandemic has been accompanied by a burst of modeling activity, and a renewed interest in the estimation of model parameters of the dynamical system describing the epidemics. A literature review of the parameter estimation and inverse problems relevant for our discussion is included in the subsequent discussion. In the present article, we propose a Bayesian particle filtering algorithm for estimating dynamically the state of the epidemic and the related model parameters based on daily counts of new infected cases. The estimation employs an underlying local SEIR-type model, which has been modified to address the characteristic features of the COVID-19 contagion.

Currently, a wealth of testing data are available, however, the data are highly non-uniform both spatially and temporally. In the beginning of the epidemics, in particular, COVID testing was largely restricted to individuals displaying the COVID-19 symptoms, and this continues to be the case with available data from many geographical regions. Therefore, the inverse problem of estimating the state and the model parameters is formulated here in terms of the symptomatic infection count data only. Observe that even if the model parameters were known, the knowledge of the size of the symptomatic infected cohort is not sufficient for prediction, as it is well understood that presymptomatic and asymptomatic individuals play a significant role in transmitting this virus. Consequently, the importance of being able to assess the contributions of untested asymptomatic cases to the epidemic dynamics has been acknowledged in COVID-19 research. Early in the pandemic outbreak, studies in the US [2] and Europe [3, 4] pointed towards a significantly high ratio of asymptomatic to symptomatic cases, while current estimates, based on more extensive testing suggest that roughly one out of three cases is asymptomatic. The particle filtering algorithm proposed in this article, together with the local dynamic model, provide a set of efficient tools to estimate this ratio from the daily count of new, mostly symptomatic, confirmed infections.

Since the groundbreaking work of MacDonald on malaria [5], one of the key indicators in epidemiology to monitor and predict the course of an epidemic is the basic reproduction number R_0, or R_t for a time dependent version of it, thoroughly discussed in the literature,

see, e.g., [6–11]. The literature on estimating the basic reproduction number of COVID-19 is vast, see, e.g., [12] for a review of some of the literature. The proposed particle filtering approach provides a means of estimating the evolution of R_t, thus to monitor the current trend in the epidemic while assessing the state of the epidemics. Characteristic to Bayesian methods, uncertainty quantification of the estimated quantities are included.

1.1. Our contribution

In this work, we propose a novel modification of the standard SEIR model, referred to as SE(A)IR model, comprising an added asymptomatic cohort and simultaneously reducing the model so that the infection data are sensitive to the size of the asymptomatic cohort, which is a key component of the inverse problem. The particle filtering approach to solve the inverse problem in the Bayesian framework shows that it is possible to stably recover the state and the parameters of the SE(A)IR model from publicly reported data. We demonstrate that the model and the corresponding inverse solver applied to real data produce estimates that are well in agreement with the current understanding of the infection dynamics and the recorded data. We also show that the ratio of asymptomatic to symptomatic infections satisfy a novel Riccati equation that can be used to define a stable equilibrium of this ratio. The equation turns out to be more than a mathematical curiosity, as numerical experiments with synthetic and real data indicate that when the epidemics is not propagating aggressively, the value of this ratio remains close to the equilibrium value and falls below it in the waning phase, thus providing a new predictive tool for the infection dynamics. Furthermore, having a way to estimate the dynamic state, we define new easily interpretable alternatives for the basic reproduction number that provide detailed information about the current trends in the dynamics of the symptomatic and asymptomatic cohorts.

The paper is organized as follows. In section 2, we introduce a parametric local compartment model for the COVID-19 epidemic and explain how it takes into account asymptomatic transmission, which is a characteristic feature of the disease. Unlike most models proposed in the literature, this compartment model is used only locally in time to describe the propagation of the state vectors over a time interval of one day, which is the frequency at which the data are updated. This way, the proposed particle filtering algorithm is not too committal on the underlying model. Moreover, possible shortcomings of the model to describe the underlying reality are continuously compensated for by a stochastic innovation process. Section 3 presents two Bayesian particle filtering algorithms for estimating the state vectors and the parameters; the details of the implementation are presented in the appendices A and B. Finally, section 4 contains computed examples based both on synthetic and real data.

2. COVID-19 epidemiology model

In this section, we describe the computational model that constitutes the forward model for the inverse problem. Unlike in standard parameter estimation approaches, the differential equation based dynamic model will be used only locally in time to describe the time evolution from one observation instance to the next, therefore it is not to be understood as to a global evolution model of the epidemics.

Compartment models [13–15] in mathematical epidemiology partition a presumably homogenous and well-mixed population into cohorts of individuals at different stages of the infection. Proposed nearly a century ago by Kermack and McKendrick [1], the popular SIR model, with separate compartments for susceptible (S), infected (I) and recovered (R), incorporates population dynamics into the previous, purely phenomenological statistical models dating back to the works of William Farr in early 19th century, which still have a life of their own

in modeling the COVID-19 epidemic [16]. Variants to the compartmental SIR model have also been developed, such as the SEIR model that include an exposed (E) compartment for individuals that have been exposed to contagion but are not infectious.

A major challenge for the control and containment of the COVID-19 epidemic is the spread of the infection by a large portion of asymptomatic or lightly symptomatic infectious individuals who are unaware of being vectors of the virus. This is especially problematic when testing priority is given to symptomatic individuals and vulnerable populations, thus the size of the asymptomatic cohort must be estimated indirectly. In the next subsection, we propose a compartment model that can be used to obtain such an estimate, and discuss its advantages and limitations.

2.1. SE(A)IR model

We begin by considering a modification of the classical SEIR model where the infected cohort I is subdivided into two infectious groups according to the manifestation of symptoms, denoting by A the asymptomatic subcohort and by I the symptomatic subcohort. Hence, while E and A are both asymptomatic and infected, the E cohort is not infectious, as opposed to A that sheds the virus and is infectious. This compartment model, schematically represented by the branching flow diagram in the left panel of figure 1, is governed by the following system of differential equations

$$\begin{aligned}
\frac{\mathrm{d}S}{\mathrm{d}t} &= -\varphi_1,\\
\frac{\mathrm{d}E}{\mathrm{d}t} &= \varphi_1 - \varphi_2 - \varphi_3^{(1)},\\
\frac{\mathrm{d}A}{\mathrm{d}t} &= \varphi_3^{(1)} - \varphi_3^{(2)},\\
\frac{\mathrm{d}I}{\mathrm{d}t} &= \varphi_2 - \varphi_4 - \varphi_5,\\
\frac{\mathrm{d}R}{\mathrm{d}t} &= \varphi_3^{(2)} + \varphi_4,
\end{aligned} \tag{1}$$

where the functional form of the fluxes will be specified below. A similar model was recently investigated, e.g., in [17, 18]. In this article, we consider COVID-19 data consisting of the daily count of newly reported symptomatic infections, corresponding to observations of the flux φ_2 in the absence of additional population-level inputs, this is the data that must be used to estimate the cohort sizes and model parameters. In particular, if no data concerning the asymptomatic cohort dynamics are available, the values of the fluxes $\varphi_3^{(1)}$ and $\varphi_3^{(2)}$ can be set rather arbitrarily, since they are weakly connected with the fluxes in the lower branch of the flow diagram, whose values are part of the observations. To estimate the size of the asymptomatic cohort in the absence of additional information, we reduce the model while retaining many of the features of the extended model that make it particularly well suited for COVID-19. The logic guiding the modification is similar to that used in metabolic network model reduction [19], when lumping enzymatic reactions whose parameters cannot be estimated from the data.

In our reduced model, the fictitious compartment $E(A)$ embeds the asymptomatic cohort into the exposed one, as depicted in the right panel of figure 1, thus the size of the new cohort is

$$E(A) = E + A,$$

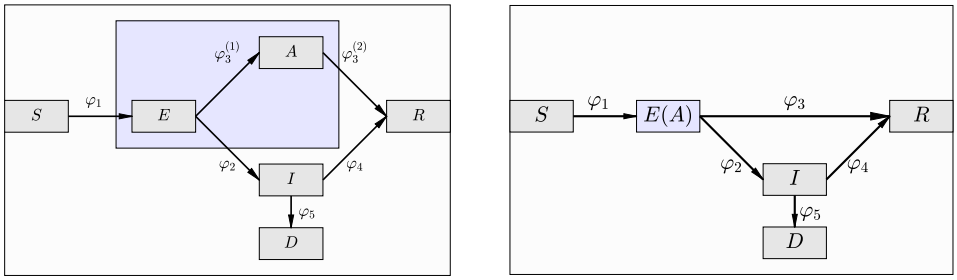

Figure 1. Left: the compartment diagram of the model that includes both symptomatic and asymptomatic infected cohorts. After a non-infectious incubation period, the exposed individuals branch either to symptomatic (I) or asymptomatic (A) infectious compartments, with unknown frequencies and unknown reasons. Right: the modified SE(A)IR model where the two compartments E and A on the blue background are lumped together to form the fictitious $E(A)$ compartment comprising both the exposed and the asymptomatic cohorts. The two fluxes $\varphi_3^{(1)}$ and $\varphi_3^{(2)}$ are merged to a single flux φ_3.

and from (1) it follows that

$$\frac{\mathrm{d}E(A)}{\mathrm{d}t} = \frac{\mathrm{d}E}{\mathrm{d}t} + \frac{\mathrm{d}A}{\mathrm{d}t} = \varphi_1 - \varphi_2 - \varphi_3^{(2)}.$$

In this manner, the flux $\varphi_3^{(1)}$, that describes a variation internal to the lumped compartment, is no longer part of the model governing equations. As in the right diagram of figure 1, we denote the flux $\varphi_3^{(2)}$ simply by φ_3. Assuming, for simplicity, that both asymptomatic and symptomatic individuals move to the recovered compartment at the same relative rate, denoted by γ, it follows that

$$\varphi_3 = \gamma E(A), \qquad \varphi_4 = \gamma I.$$

The flux φ_2 is proportional to the rate at which a portion of the exposed individuals develops symptoms. For simplicity, we assume that the flux is proportional to the size of the compounded cohort, that is,

$$\varphi_2 = \eta E(A).$$

One of the most relevant parameters to describe the spread of an epidemic is the rate at which susceptible individuals are infected. In the classical SEIR model, the transmission flux assumes the form

$$\varphi_1 = \beta \frac{I}{N_\mathrm{p}} S,$$

where β is the transmission rate and N_p is the population size. Our model accounts for the possibility that virus transmission can occur through contact with the asymptomatic cohort within $E(A)$, as is the case in COVID-19, while most infected individuals who have developed symptoms are either hospitalized or in self-isolation, thus playing a limiting role in the exposure of susceptible individuals to contagion. Acknowledging the differing roles that symptomatic and asymptomatic virus shedders play in transmission to susceptibles, we postulate

$$\varphi_1 = \beta \frac{pE(A) + qI}{N_\mathrm{p}} S,$$

where p and q are frequencies, $0 \leqslant p, q \leqslant 1$. The frequency p may be interpreted as the fraction of individuals in the compound cohort $E(A)$ who are infectious, while q reflects the limited exposure to self-isolated or hospitalized symptomatic individuals as compared to the asymptomatic ones. Classical SEIR models, where usually it is assumed that the exposed cohort is not infectious, can be obtained by setting $p = 0$, $q = 1$. When using our model to understand COVID-19 dynamics, we assume that $p > q$. The transmission rate β, which is the quantity of primary interest when it comes to assessing how fast epidemics spread, integrates elements related to the characteristics of the pathogen determining the probability of infection of a given susceptible contact, and factors related to social behavior, including the number and nature of daily contacts. In summary, the governing equations of the proposed model are

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\beta \frac{pE(A) + qI}{N_{\mathrm{p}}} S, \tag{2}$$

$$\frac{\mathrm{d}E(A)}{\mathrm{d}t} = \beta \frac{pE(A) + qI}{N_{\mathrm{p}}} S - \eta E(A) - \gamma E(A), \tag{3}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \eta E(A) - \gamma I - \mu I, \tag{4}$$

$$\frac{\mathrm{d}R}{\mathrm{d}t} = \gamma E(A) + \gamma I, \tag{5}$$

where β is the transmission rate, γ is the recovery rate, η is the rate at which symptoms develop that integrates the incubation process $E \to I$, and μ is the death rate.

As pointed out, the transmission rate β reflects not only properties of the pathogen, but also factors related to human behavior that change in the course of an epidemic, therefore in the filtering approach that we propose below to estimate the course of the epidemic, β is modeled as a time dependent parameter. In the absence of a known deterministic dynamical evolution model for β, its changes will be described in terms of a stochastic geometric random walk.

Before describing the computational methodology at the heart of our model predictions, some qualitative comments about the model are in order. In classical SEIR models, the presence of the exposed cohort E adds a time delay corresponding to the incubation period that could not be accounted for in SIR models. While our expression for φ_1 implicitly assumes that the infected and asymptomatic individuals are immediately infectious, the flux φ_2 introduces a slight delay in changes to the I cohort when the transmission rate changes. Moreover, if $p \neq 0$, we may write

$$\varphi_1 = \beta p \frac{E(A) + (q/p)I}{N_{\mathrm{p}}} S,$$

and redefining β so as to comprise scaling by p, the model can be written in terms of the ratio q/p. Hence, without a loss of generality, we may set $p = 1$, and assume $q < 1$. In our numerical examples we set $q = 0.1$. A discussion of how the value of q affects the estimated quantities is presented in the context of the computed results. We remark that the assumption that the symptomatic and asymptomatic individuals recover at the same rate is not essential for the methodology developed below, and can be easily removed.

The merging of the compartment E and A into the $E(A)$ compartment is in agreement with the observation that the transition from exposed non-infective (E) to asymptomatic infective (A) is not fully understood for COVID-19, as it has been observed that the virus load increases gradually after contagion, regardless of whether symptoms develop. The choice of having

separate compartments for symptomatic and asymptomatic cohorts follows from the observation that it is the onset of symptoms, rather that the increase in viral load, that affects the behavior of the infected individual (self-isolation, hospitalization) and dramatically changes their role in the spread of the infection, as reflected in our model.

2.2. The forward model and the inverse problem

The data that are used to inform our parameterized model, as well as to update the estimation of the size of the cohorts over time comprise the number of confirmed, symptomatic new daily cases. Assuming that the time is given in units of one day, the observed data are noisy realizations of the flux,

$$\int_{t-1}^{t} \varphi_2(t')\mathrm{d}t' \approx \varphi_2(t) = \eta E(A)(t), \quad t = 1, 2, \ldots$$

The daily case count constituting the data is an integer, and we model it as a Poisson distributed random variable,

$$B_t \sim \text{Poisson}(\varphi_2(t)), \quad t = 1, 2, \ldots$$

Denoting by z_t the state vector, $z_t = (S_t, E(A)_t, I_t, R_t)$ and by (β_t, η, γ) the model parameters, the forward model can be described as

$$(z_0, \{\beta_t\}_{t=1}^T, \eta, \gamma) \to \{z_t\}_{t=1}^T \to \{\varphi_2(t)\}_{t=1}^T \to \{B_t\}_{t=1}^T,$$

for some $T > 0$. Here the mortality rate is assumed to be fixed and known. The inverse problem of estimating the size of the cohorts and the parameter values from the observations is solved here in a sequential manner, estimating simultaneously z_{t+1} and β_{t+1} using the previous estimates z_t and β_t together with the new data B_{t+1}.

The sequential estimation process, based on a Bayesian statistical model described in detail in the following section, has a couple of features worth highlighting. First, the estimation process produces a distribution of values z_t and β_t, allowing us to compute derived quantities and to obtain credibility estimates for these quantities. Second, the estimation process assumes that the state vector satisfies the model (2)–(5) only approximately, acknowledging that a simplified well-mixed closed dynamical model cannot represent accurately the full complexity of a heterogeneous population which is in interaction with the surrounding world. The implicit integration of the model discrepancy [20] into the computational scheme makes the approach more flexible than those based on a deterministic forward model, and consequently less prone to be biased by model assumptions.

Since the outbreak of COVID-19, there has been a significant amount of literature related to compartment models and parameter estimation. Global optimization algorithms for estimating the parameters of a SIR model from infection data were discussed in [21, 22]. An analytical method based on special function representations was discussed in [23], and a cluster analysis was proposed in [24]. Time dependent parameter functions in the SIR model were estimated in [25] using Tikhonov regularization and a Levenberg–Marquardt approach, and similar approaches were proposed and discussed in [26, 27]. More complex SEIR-like model and further extensions were proposed in [28], where the inverse problem was solved using stochastic optimization methods such as simulated annealing. Similarly, a SEIR model was used in [18], and its parameters were estimated using particle swarm optimization. A number of methods based on the Bayesian paradigm for solving inverse problems have been proposed: in [29], the early infection data from China was analyzed using a SEIR model using sequential Monte Carlo methods, and variational Bayesian methods with a SIR model were used in

[30]. A more complex SEIR-based model (SAPHIRE) was proposed in [31], and Markov chain Monte Carlo (MCMC) methods were used to estimate the parameters. The article [32], which is methodologically closer to ours, proposes a stochastic SEIR model and uses ensemble Kalman filtering to estimate the model parameters. For other contributions using Bayesian techniques with different modeling approaches, see, e.g., [33, 34].

In the discussion to follow, we simplify the notation by using E, instead of $E(A)$, to denote the exposed/asymptomatic cohort size.

3. Bayesian particle filtering

Bayesian filtering methods estimate the time evolution of state vectors of a dynamic model from limited observations in a sequential manner. In keeping with the Bayesian paradigm for inverse problems, all unknown quantities, hence the state vector and possibly any unknown parameters, are modeled as random variables. The solution computed via Bayesian filtering is the time evolving posterior density of the state vector and parameters, while the evolution model can be viewed as part of the prior. We use the convention of denoting random variables with upper case letters and their realizations with the corresponding lower case letters. We refer to [35–37] for an overview of particle filtering, and to [38–40] for the particular type of applications to ODE systems. For a recent review of particle filters in the context of data assimilation, see [41].

Let $\{X_t\}$ denote a discrete time Markov process, $t = 0, 1, 2, \ldots$, the time step being one day, with the transition probability distribution $\pi_{X_{t+1}|X_t}(x_{t+1}|x_t)$. Furthermore, let $\{B_t\}$ denote the stochastic process representing the observations, and let $\pi_{B_t|X_t}(b_t|x_t)$ denote the likelihood density, where we implicitly assume that the current observation B_t depends on the past only through the current state X_t. Finally, let $\mathscr{B}_t$ denote the cumulative data up to time t, that is,

$$\mathscr{B}_t = \{B_1, B_2, \ldots, B_t\},$$

and by $\mathcal{B}_t$ the set of observed realizations, $\mathfrak{B}_t = \{b_1, b_2, \ldots, b_t\}$. In Bayesian filtering algorithms the update of the posterior distribution from time t to time $t+1$ is carried out in two consecutive steps,

$$\pi_{X_t|\mathscr{B}_t} \to \pi_{X_{t+1}|\mathscr{B}_t} \to \pi_{X_{t+1}|\mathscr{B}_{t+1}},$$

where the first step is referred to as the propagation, or prediction step and the second as the analysis, or correction step. The propagation step is accomplished through the Chapman–Kolmogorov formula,

$$\pi_{X_{t+1}|\mathscr{B}_t}(x_{t+1}|\mathfrak{B}_t) = \int \pi_{X_{t+1}|X_t}(x_{t+1}|x_t)\pi_{X_t|\mathscr{B}_t}(x_t|\mathfrak{B}_t)\mathrm{d}x_t,$$

while the analysis step builds on Bayes' formula,

$$\pi_{X_{t+1}|\mathscr{B}_{t+1}}(x_{t+1}|\mathfrak{B}_{t+1}) = \pi_{X_{t+1}|\mathscr{B}_t}(x_{t+1}|\mathfrak{B}_t)\pi_{B_{t+1}|X_{t+1}}(b_{t+1}|x_{t+1}),$$

with the predicted distribution of X_{t+1} acting as the prior when the next observation arrives. We now proceed to specify the transition probability kernel and the likelihood for our model, and describe the computational steps for the numerical implementation of the Bayesian filter. We then extend the discussion to include the estimation of static parameters.

3.1. Particle filtering with fixed static parameters

Consider the SE(A)IR model introduced in the previous section, and let

$$z_t = \begin{bmatrix} S_t \\ E_t \\ I_t \\ R_t \end{bmatrix}, \quad t = 1, 2, \ldots$$

be the corresponding state vector at time t. We assume that the infectivity parameter β may vary over time and denote its value at time t by β_t. We collect the other model parameters in the vector $\theta = (\gamma, \eta, \mu)$, which is assumed to be static and, for the time being, known.

We denote formally by ψ the numerical propagator advancing the state variable from one day to the next, $t \to t+1$,

$$\psi(z_t, \beta_t) = z_{t+1};$$

with the dependency on the static parameters suppressed because they remain constant in time. In our computations, the time integration of the system (2)–(5) is performed by means of a standard ODE solver such as Runge–Kutta, keeping the value of the infectivity parameter β_t fixed during the one day propagation step.

Formally, we write a propagation model of the form

$$x_t = \begin{bmatrix} z_t \\ \beta_t \end{bmatrix} \to F(x_t) = \begin{bmatrix} \psi(z_t, \beta_t) \\ \beta_t \end{bmatrix} = \widehat{x}_{t+1}, \tag{6}$$

and we account for uncertainties both in the state vector z_t and the parameter vector β_t by introducing an innovation term. We guarantee that all components of the state vector and the parameter β remain nonnegative by introducing a multiplicative innovation that follows a geometric random walk model,

$$\log X_{t+1} \sim \mathcal{N}(\log \widehat{x}_{t+1}, \mathsf{C}), \tag{7}$$

where $\mathsf{C} \in \mathbb{R}^4$ is a positive definite diagonal matrix. Thus the transition probability density $\pi_{X_{t+1}|X_t}(x_{t+1}|x_t)$ is defined though (6) and (7).

In the definition of the likelihood, we assume that the data consist of realizations b_t of the daily count of new infections, denoted by B_t. This count is assumed to be Poisson distributed, with the expectation equal to the flux $\varphi_2(t)$, that is

$$B_t \sim \text{Poisson}(\varphi_2(t)), \quad \text{where } \varphi_2(t) = \eta E_t.$$

Therefore, the likelihood of the observed number b_t of new infections is

$$\pi_{B_t|Z_t}(b_t|z_t) = \frac{(\eta E_t)^{b_t}}{b_t!} \mathrm{e}^{-\eta E_t}.$$

To initialize the process, we need to specify the value of the state vector at the time $t = 0$ prior to observing the first infections. If the initial state is unknown, it becomes part of the estimation problem also. We postulate that before the first infection is observed, there is an unknown number of asymptomatic and symptomatic individuals in the community and that the initial number of asymptomatic cases follows a Poisson distribution with a uniformly distributed expected value $\Lambda_E, \Lambda_I \sim \text{Uniform}([0, \lambda_{\max}])$. Moreover, we assume that β_0 follows a uniform distribution over some interval $[\beta_{\min}, \beta_{\max}]$.

We are now ready to outline the particle filter (PF) algorithm for estimating the state vector X_t based on the daily new infection count. Assume that at time t a sample $\{x_t^1, x_t^2, \ldots, x_t^N\}$ drawn from the distribution $\pi_{X_t|\mathscr{B}_t}$ is available, and that w_t^j is the weight associated with the sample point x_t^j, so that we can approximate the underlying density by

$$\pi_{X_t|\mathcal{B}_t}(x_t|\mathfrak{B}_t) \approx \sum_{j=1}^{N} w_t^j \delta_{x_t^j}(x_t).$$

Substituting this particle approximation of the density in the Chapman–Kolmogorov formula gives

$$\begin{aligned}\pi_{X_{t+1}|\mathscr{B}_t}(x_{t+1}|\mathfrak{B}_t) &= \int \pi_{X_{t+1}|X_t}(x_{t+1}|x_t)\pi_{X_t|\mathscr{B}_t}(x_t|\mathfrak{B}_t)\mathrm{d}x_t \\ &\approx \sum_{j=1}^{N} w_t^j \pi_{X_{t+1}|X_t}(x_{t+1}|x_t^j), \end{aligned} \tag{8}$$

and subsequently, substituting in Bayes' formula, we obtain the updating formula

$$\pi_{X_{t+1}|\mathscr{B}_{t+1}}(x_{t+1}|\mathfrak{B}_{t+1}) \approx \sum_{j=1}^{N} w_t^j \pi_{X_{t+1}|X_t}(x_{t+1}|x_t^j)\pi_{B_{t+1}|X_{t+1}}(b_{t+1}|x_{t+1}). \tag{9}$$

Let $\widehat{x}_{t+1}^j$ denote a propagated predictor particle associated with x_t^j, that is,

$$\widehat{x}_{t+1}^j = F(x_t^j),$$

where F is defined in (6). At the arrival of the next observation b_{t+1}, the likelihood $\pi_{B_{t+1}|X_{t+1}}(b_{t+1}|\widehat{x}_{t+1}^j)$ expresses how good the predictor is at explaining the data and can be used to reformulate (9) as

$$\begin{aligned}\pi_{X_{t+1}|\mathscr{B}_{t+1}}(x_{t+1}|\mathcal{B}_{t+1}) \approx \sum_{j=1}^{N} &\underbrace{\left\{ w_t^j \pi_{B_{t+1}|X_{t+1}}(b_{t+1}|\widehat{x}_{t+1}^j) \right\}}_{\text{(a)}} \\ &\times \underbrace{\left\{ \frac{\pi_{B_{t+1}|X_{t+1}}(b_{t+1}|x_{t+1})}{\pi_{B_{t+1}|X_{t+1}}(b_{t+1}|\widehat{x}_{t+1}^j)} \right\}}_{\text{(b)}} \underbrace{\pi_{X_{t+1}|X_t}(x_{t+1}|x_t^j)}_{\text{(c)}}. \end{aligned} \tag{10}$$

The three factors whose product accounts for each particle contribution have natural interpretations in terms of the fitness of the particle at explaining the data. The factor (a) expresses the relevance of the particle $\widehat{x}_{t+1}^j$ in explaining the new data, by combining the importance of its predecessor, encoded in w_t^j, and the likelihood of the observed data point; the transition kernel (c) accounts for the fact that the next generation particle is generated from its predictor according to (7); the factor (b) weighs the importance of the new particle relative to the predictor. This hierarchical organization is the backbone of the particle filter algorithm. For completeness, we present the details of the implementation from (10) in appendix A.

Particle filtering algorithm with known static parameters

Initialize: draw N independent realizations of $\beta_0 \sim \text{Uniform}([\beta_{\min}, \beta_{\max}])$ and $\Lambda_E, \Lambda_I \sim \text{Uniform}([0, \lambda_{\max}])$,

$$\{\beta_0^1, \beta_0^2, \ldots, \beta_0^N\}, \quad \{\lambda_E^1, \lambda_E^2, \ldots, \lambda_E^N\}, \quad \{\lambda_I^1, \lambda_I^2, \ldots, \lambda_I^N\},$$

generate N realizations of the initial state Z_0,

$$z_0^j = \begin{bmatrix} N - \lambda_E^j - \lambda_I^j \\ \lambda_E^j \\ \lambda_I^j \\ 0 \end{bmatrix}, \quad j = 1, 2, \ldots, N,$$

and define the initial particle cloud

$$\{x_0^1, x_0^2, \ldots, x_0^N\}, \quad x_0^j = \begin{bmatrix} z_0^j \\ \beta_0^j \end{bmatrix}.$$

Set $t = 0$.

While $t < t_{\max}$ **do**

(a) Propagate the particles according to (6) to generate the predictive particle cloud

$$\{\widehat{x}_{t+1}^1, \widehat{x}_{t+1}^2, \ldots, \widehat{x}_{t+1}^N\}.$$

(b) Extract the second component $\widehat{E}_{t+1}^j$ of each propagated particle, and compute the weights,

$$\widehat{g}_{t+1}^j = w_t^j \frac{(\eta \widehat{E}_{t+1}^j)^{b_{t+1}}}{b_{t+1}!} \mathrm{e}^{-\eta \widehat{E}_{t+1}^j}, \quad \widehat{g}_{t+1}^j \leftarrow \frac{\widehat{g}_{t+1}^j}{\sum_\ell \widehat{g}_{t+1}^\ell}.$$

(c) Sample with replacement N indices $\ell_j \in \{1, 2, \ldots, N\}$, $j = 1, 2, \ldots, N$, using the normalized weights g_{t+1}^j as probabilities. Define the new resampled predictive cloud,

$$\widehat{x}_{t+1}^j \leftarrow \widehat{x}_{t+1}^{\ell_j}.$$

Generate a new particle cloud through the innovation process,

$$\log x_{t+1}^j = \log \widehat{x}_{t+1}^j + \mathsf{C}^{1/2} w^j, \quad w^j \sim \mathcal{N}(0, \mathsf{I}_5).$$

(d) Extract the second component E_{t+1}^j from each new particle, evaluate the normalized likelihood,

$$g_{t+1}^j = \frac{(\eta E_{t+1}^j)^{b_{t+1}}}{b_{t+1}!} \mathrm{e}^{-\eta E_{t+1}^j}, \quad g_{t+1}^j \leftarrow \frac{g_{t+1}^j}{\sum_\ell g_{t+1}^\ell}.$$

and update the weights,

$$w_{t+1}^j = \frac{g_{t+1}^j}{\widehat{g}_{t+1}^j}.$$

Advance the counter $t \to t + 1$.

end do

In this version of the PF algorithm, the model parameters γ, η and μ are assumed to be fixed and known. It is also possible to modify the algorithm to estimate these parameters. In that case, to estimate the death rate μ, the information about the deceased must be included in the data, as the new infection count is essentially insensitive to that parameter. The parameters γ and η are less time-dependent than the infectivity β, and we set their values according to what is suggested in the literature.

3.2. Particle filter with unknown static parameters

The algorithm outlined in the previous section assumes that the static parameters γ, η and μ are fixed and known. A coarse estimate of these parameters can be obtained using the average recovery time $T_{\rm rec}$ as well as the estimated time $T_{\rm sym}$ from exposure to onset of symptoms. However, these times are ill-defined and difficult to observe directly, with the variability between individuals contributing to the uncertainty. Therefore, we outline an extension of the PF algorithm to also estimate the static parameters that is an adaptation of the particle filter/sequential Monte Carlo (PF-SMC) algorithm discussed in [35, 39]. In this version, the vector x_t in (6) is augmented by the parameter vector $\theta_t = (\gamma_t, \eta_t)$, whose components are assumed, *a priori*, to be constants or almost constants, but unknowns. To prevent the algorithm for compensating for changes in the time dependent β_t erroneously by varying significantly the values of γ_t and η_t, we define a propagation model of the form

$$x_t = \begin{bmatrix} z_t \\ \beta_t \end{bmatrix} \to F(x_t) = \begin{bmatrix} \psi(z_t, \beta_t, \theta_t) \\ \beta_t \mathrm{e}^{\alpha w_t} \end{bmatrix} = \widehat{x}_{t+1}, \qquad \widehat{\theta}_{t+1} = \theta_t, \tag{11}$$

where $\alpha > 0$ is a parameter controlling the variability of β_t from one time step to the next, and $w_t \sim \mathcal{N}(0, 1)$, that is, the propagation model has a random component.

The innovation process for x_t uses a geometric random walk model similar to (7). However, to avoid artificial diffusion of the static parameters, the component θ_t is treated slightly differently. Given the predictive sample $\widehat{\theta}^1_{t+1}, \ldots, \widehat{\theta}^N_{t+1}$, we calculate the mean and the covariance of the logarithms of the particles,

$$\overline{\xi}_{t+1} = \sum_{j=1}^N w_t^j \log \widehat{\theta}^j_{t+1}, \qquad \Sigma_{t+1} = \sum_{j=1}^N w_t^j (\log \widehat{\theta}^j_{t+1} - \overline{\xi}_{t+1})(\log \widehat{\theta}^j_{t+1} - \overline{\xi}_{t+1})^{\mathsf{T}}.$$

Let $0 < h < 1$ be a parameter controlling the variance of the innovation, and $a = \sqrt{1-h^2}$. With these notations, we introduce the innovation model

$$\log \Theta_{t+1} \sim \mathcal{N}(a \log \theta_t + (1-a)\overline{\xi}_{t+1}, h^2 \Sigma_{t+1}).$$

For completeness and for convenience of the reader, a detailed justification of this formula is provided in appendix B.

We are now ready to summarize the steps in algorithmic form.

Particle filtering algorithm with unknown static parameters

Initialize: draw N independent realizations of $\beta_0 \sim \text{Uniform}([\beta_{\min}, \beta_{\max}])$, $\gamma_0 \sim \text{Uniform}([\gamma_{\min}, \gamma_{\max}])$, $\eta_0 \sim \text{Uniform}([\eta_{\min}, \eta_{\max}])$ and $\Lambda_E, \Lambda_I \sim \text{Uniform}([0, \lambda_{\max}])$,

$$\{\beta_0^1, \beta_0^2, \ldots, \beta_0^N\}, \quad \{\gamma_0^1, \gamma_0^2, \ldots, \gamma_0^N\}, \quad \{\eta_0^1, \eta_0^2, \ldots, \eta_0^N\}, \quad \{\lambda_E^1, \lambda_E^2, \ldots, \lambda_E^N\}, \quad \{\lambda_I^1, \lambda_I^2, \ldots, \lambda_I^N\},$$

generate N realizations of the initial state of Z_0,

$$z_0^j = \begin{bmatrix} N - \lambda_E^j - \lambda_I^j \\ \lambda_E^j \\ \lambda_I^j \\ 0 \end{bmatrix}, \quad j = 1, 2, \ldots, N,$$

and define the initial cloud of the particles

(*continued on next page*)

$$\{x_0^1, x_0^2, \ldots, x_0^N\}, \qquad x_0^j = \begin{bmatrix} z_0^j \\ \beta_0^j \end{bmatrix}, \qquad \theta_0^j = \begin{bmatrix} \gamma_0^j \\ \eta_0^j \end{bmatrix}.$$

Set $t = 0$.
While $\boldsymbol{t} < \boldsymbol{t}_{\max}$ do

(a) Propagate the particles according to (11) to generate the predictive particle cloud

$$\{\widehat{x}_{t+1}^1, \widehat{x}_{t+1}^2, \ldots, \widehat{x}_{t+1}^N\}, \qquad \{\widehat{\theta}_{t+1}^1, \widehat{\theta}_{t+1}^2, \ldots, \widehat{\theta}_{t+1}^N\}.$$

(b) Extract the second component $\widehat{E}_{t+1}^j$ of each of the propagated particles $\widehat{x}_{t+1}^j$, and compute the weights

$$\widehat{g}_{t+1}^j = w_t^j \frac{(\eta \widehat{E}_{t+1}^j)^{b_{t+1}}}{b_{t+1}!} \mathrm{e}^{-\eta \widehat{E}_{t+1}^j}, \quad \widehat{g}_{t+1}^j \leftarrow \frac{\widehat{g}_{t+1}^j}{\sum_\ell \widehat{g}_{t+1}^\ell}.$$

(c) Sample with replacement N indices $\ell_j \in \{1, 2, \ldots, N\}$, $j = 1, 2, \ldots, N$, with the probability weights g_{t+1}^j. Define the new resampled predictive cloud

$$\widehat{x}_{t+1}^j \leftarrow \widehat{x}_{t+1}^{\ell_j} \qquad \widehat{\theta}_{t+1}^j \leftarrow \widehat{\theta}_{t+1}^{\ell_j}.$$

Generate a new particle cloud through the innovation process

$$\log x_{t+1}^j = \log \widehat{x}_{t+1}^j + \mathsf{C}^{1/2} w^j, \quad w^j \sim \mathcal{N}(0, \mathsf{I}_5),$$
$$\log \theta_{t+1}^j = a \log \widehat{\theta}_{t+1}^j + (1-a)\overline{\xi}_{t+1} + \Sigma^{1/2} \omega^j, \quad \omega^j \sim \mathcal{N}(0, \mathsf{I}_2),$$

(d) Extract the second component E_{t+1}^j from each new particle x_{t+1}^j, and evaluate the normalized likelihood

$$g_{t+1}^j = \frac{(\eta E_{t+1}^j)^{b_{t+1}}}{b_{t+1}!} \mathrm{e}^{-\eta E_{t+1}^j}, \quad g_{t+1}^j \leftarrow \frac{g_{t+1}^j}{\sum_\ell g_{t+1}^\ell}.$$

Update the weights,

$$w_{t+1}^j = \frac{g_{t+1}^j}{\widehat{g}_{t+1}^j}.$$

Advance the counter $t \to t+1$.

end do

We close this section by noting that the proposed algorithm is slightly simpler than the one in [39], where an additional repropagation step was included. Numerical test indicate that in the present framework, there are no notable difference in results with and without the repropagation step.

3.3. Basic reproduction numbers

When assessing and predicting the evolution of an epidemic, the estimated parameters have the drawback that they are meaningful only in the framework of the model, while derived dimensionless parameters, or Π-numbers of the model [15] may be more informative, as they have a wider interpretation. The most significant Π-number for epidemiological models is the basic reproduction number R_t. The standard definition of R_t is based on the concept of the *next generation matrix*, see [6–11]. In addition to the classical R_t, we will consider below two dimensionless quantities that are directly related to the derivatives of the compartments E and I of the model, referred to as disease compartments in the discussion about reproduction number.

3.3.1. Next generation matrix based R_t. The first step in the derivation of the formula for the standard basic reproduction number is to write equations (3) and (4) describing the dynamics of the infection-carrying, or disease compartments as

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} E \\ I \end{bmatrix} = \underbrace{\begin{bmatrix} \beta p\nu_{\mathrm{S}} & \beta q\nu_{\mathrm{S}} \\ 0 & 0 \end{bmatrix}}_{\mathsf{F}} \begin{bmatrix} E \\ I \end{bmatrix} - \underbrace{\begin{bmatrix} \eta+\gamma & 0 \\ -\eta & \gamma+\mu \end{bmatrix}}_{\mathsf{V}} \begin{bmatrix} E \\ I \end{bmatrix} = (\mathsf{F}-\mathsf{V}) \begin{bmatrix} E \\ I \end{bmatrix},$$

where $\nu_{\mathrm{S}} = \nu_{\mathrm{S}}(t)$ denotes the frequency S/N_{p} of the susceptible cohort. The matrix F describes the flux into the disease compartments E and I, and V the clearance of the disease compartments. The next generation matrix G is defined as

$$\mathsf{G} = \mathsf{F}\mathsf{V}^{-1} = \begin{bmatrix} \left(p + \dfrac{q\eta}{\gamma+\mu}\right)\dfrac{\beta\nu_{\mathrm{S}}}{\gamma+\eta} & \dfrac{q}{\gamma+\mu}\beta \\ 0 & 0 \end{bmatrix},$$

see, e.g. [11] for further explanations. The basic reproduction number is defined as the spectral radius of the next generation matrix, thus for the current model

$$R_t^{\mathrm{NG}} = \left(p + \frac{q\eta}{\gamma+\mu}\right)\frac{\beta\nu_{\mathrm{S}}}{\gamma+\eta}, \tag{12}$$

where the superscript indicates that the derivation is based on the next generation matrix.

3.3.2. Control of infected and exposed compartments. The basic reproduction number R_t^{NG} defined above relies on the linearization around a hypothetical disease-free state. Alternative tools to analyze a spreading epidemic can be related to the instantaneous growth rates of the disease compartments at the current state of the epidemic, without a reference to a disease-free state. To this end, we denote by $\rho = \rho(t)$ the ratio of the exposed/asymptomatic and infected cohort sizes,

$$\rho = \frac{E}{I}. \tag{13}$$

Combining (3) and (4) we obtain

$$\begin{aligned} \frac{\mathrm{d}E}{\mathrm{d}t} &= \beta(pE+qI)\nu_{\mathrm{S}} - (\gamma+\eta)E = -(\gamma+\eta)\left(1 - \frac{\beta(p+q\rho^{-1})}{\gamma+\eta}\right)E \\ &= -(\gamma+\eta)\left(1-R_t^E\right)E, \end{aligned}$$

and, likewise,

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \eta E - (\gamma+\mu)I = -(\gamma+\mu)\left(1 - \frac{\eta\rho}{\gamma+\mu}\right)I = -(\gamma+\mu)\left(1-R_t^I\right)I,$$

indicating that, similarly to the classical basic reproduction number for the SIR model, the dimensionless quantities

$$R_t^E = \frac{\beta(p+q\rho^{-1})}{\gamma+\eta}, \qquad R_t^I = \frac{\eta\rho}{\gamma+\mu}, \tag{14}$$

describe whether the corresponding diseased cohorts are increasing or decreasing, the critical value being $R_t^E = R_t^I = 1$. Observe that unlike the classical basic reproduction number, these quantities do not define naturally a critical community size of susceptible individuals.

We point out that to estimate (14) from the daily infection data, it is necessary to estimate the cohort sizes, a task for which the proposed PF is particularly well suited. Moreover, for monitoring and controlling the spread of an epidemic where asymptomatic individuals may be the main vectors of contagion, as is the case in COVID-19, the ratio (13) is a quantity of interest and worthy of a closer look.

3.4. Ratio of asymptomatic to symptomatic cohorts: a closer look

Differentiating ρ with respect to time,

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = \frac{1}{I}\frac{\mathrm{d}E}{\mathrm{d}t} - \frac{E}{I^2}\frac{\mathrm{d}I}{\mathrm{d}t},$$

expressing the derivatives of E and I in terms of the governing equations (3) and (4), and simplifying, we find that ρ satisfies the Riccati equation

$$\begin{aligned}\frac{\mathrm{d}\rho}{\mathrm{d}t} &= \beta\frac{p+q\rho}{N_{\mathrm{p}}}S - (\eta+\gamma)\rho - \eta\rho^2 - (\gamma+\mu)\rho \\ &= -\eta\left(\rho^2 + \left(\frac{\eta+2\gamma+\mu}{\eta} - \frac{\beta}{\eta}q\nu_{\mathrm{S}}\right)\rho - \frac{\beta}{\eta}p\nu_{\mathrm{S}}\right), \qquad \nu_{\mathrm{S}} = \frac{S}{N_{\mathrm{p}}}. \end{aligned} \tag{15}$$

The ratio of the sizes of the asymptomatic and symptomatic cohorts has been a topic of intense discussion from the beginning of the COVID-19 pandemic, with hypothetical values ranging from as high as 50–80 (see, e.g., [2]) at the onset of the pandemic, to the current C.D.C. estimate of around 0.5. For this reason, it is of interest to investigate if the proposed model suggests the existence of an equilibrium value ρ^* for the ratio. The value of ρ at which its derivative vanishes can be expressed in closed form as

$$\rho^* = -\frac{1}{2}\left(\frac{\eta+2\gamma+\mu}{\eta} - \frac{\beta}{\eta}q\nu_{\mathrm{S}}\right) + \sqrt{\frac{1}{4}\left(\frac{\eta+2\gamma+\mu}{\eta} - \frac{\beta}{\eta}q\nu_{\mathrm{S}}\right)^2 + \frac{\beta}{\eta}p\nu_{\mathrm{S}}}, \tag{16}$$

Moreover, since the right-hand side of the equation (15) switches its sign from positive to negative as ρ is increased past the critical value ρ^*, this is a stable equilibrium. In the section of computed examples, we will discuss the significance of the equilibrium value in the light of the state of the epidemic.

4. Computed examples

We test the performance of the proposed inversion framework with both synthetic and real data.

4.1. Testing with synthetic data

Example 1. In the first experiment, three sets of simulated data are generated with the same model (2)–(5) that the PF algorithm is based on, under the assumption that the parameter β is time dependent, while the parameters γ and η are constants. The functional form the parameter β in the generative model,

$$\beta(t) = \begin{cases} \beta_0, & \text{for } 0 \leqslant t \leqslant T, \\ \beta_\infty + (\beta_0 - \beta_\infty)\mathrm{e}^{-(t-T)/\tau}, & \text{for } t > T, \end{cases} \tag{17}$$

Table 1. The parameter values used in the particle filter algorithm with fixed static parameters.

Particle filter parameters		
a priori lower bound transmission rate [1/days]	$\beta_{\min}$	0.1
a priori upper bound for transmission rate	$\beta_{\max}$	0.8
a priori upper bound of initial infectious individuals	$\lambda_{\max}$	5
Number of particles	N	20 000
Standard deviation of innovation of the state	σ_1	0.01
Standard deviation of innovation of the transmission rate	σ_2	0.1
Static model parameters		
Recovery rate	γ	1/14
Symptomatic rate	η	1/7
Death rate	μ	0.004

where $\beta_0 = 0.4$, $\beta_\infty = 0.1$, $T = 20$ d and $\tau = 5$ d, remains unchanged in all simulated data sets. The time interval of the simulation is 120 d. In the three test protocols (A)–(C), the values of the parameters γ and η used to generate the data were set to, respectively,

$$\text{(A)}: \begin{cases} \gamma = 1/14, \\ \eta = 1/7, \end{cases} \qquad \text{(B)}: \begin{cases} \gamma = 1/10, \\ \eta = 1/5, \end{cases} \qquad \text{(C)}: \begin{cases} \gamma = 1/20, \\ \eta = 1/10. \end{cases}$$

The death rate is set to $\mu = 0.004$.

The values of the model parameters and of the prior densities, as described in the algorithm in section 3, are listed in table 1. In this set of tests, the static parameters are held fixed in the PF algorithm, and they are set at the values used in protocol (A). The number of particles is $N = 20\,000$, and the initial sample for β was drawn from

$$\beta_0 \sim \text{Uniform}([\beta_{\min}, \beta_{\max}]),$$

with upper and lower bounds as listed in table 1. The total population size in the simulation is set to $N_{\text{p}} = 100\,000$, which is supposed to be approximately known, and the initial values (S_0, E_0, I_0, R_0) for the state vectors are drawn from the distribution

$$I_0, E_0 \sim \text{Uniform}([0, \lambda_{\max}]), \qquad S_0 = N_{\text{p}} - I_0 - E_0, \quad R_0 = 0,$$

see table 1 for the numerical value. The innovation covariance matrix C is a diagonal matrix of the form

$$\mathsf{C} = \text{diag}(\sigma_1^2/N_{\text{p}}^2, \sigma_1^2, \sigma_1^2, \sigma_1^2, \sigma_2^2),$$

where σ_1^2 and σ_2^2 are the variances of the uncertainty in the logarithm of the state vector Z_t and in the logarithm of the transmission rate β_t, respectively. We point out that the population size N_{p} is assumed to be only approximately known, and it is not enforced. Observe further that for the susceptible population, the variance is weighted by the population squared to keep the innovation for this cohort from becoming excessive. We have

$$S_{t+1}^j = \widehat{S}_{t+1}^j \exp\left(\frac{\sigma_1}{N_{\text{p}}} w_{t+1}^j\right) \approx \widehat{S}_{t+1}^j + \sigma_1 \frac{\widehat{S}_{t+1}^j}{N_{\text{p}}} w_{t+1}^j,$$

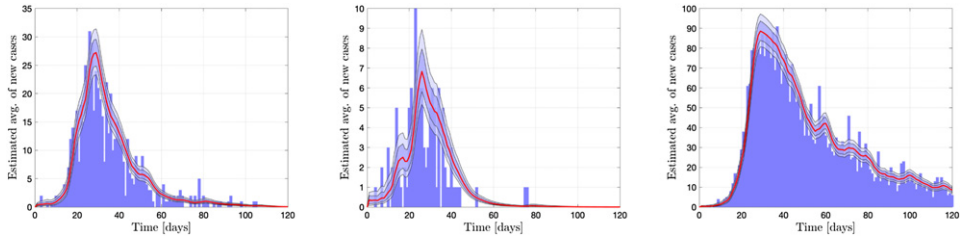

Figure 2. Simulated data corresponding to the protocols A (left), B (center) and C (right), shown as bar plots. The darker and lighter envelopes indicate the 50% and 75% estimated credible intervals of the mean of the new cases $\varphi_2 = \eta E$, and the red curve is the median value.

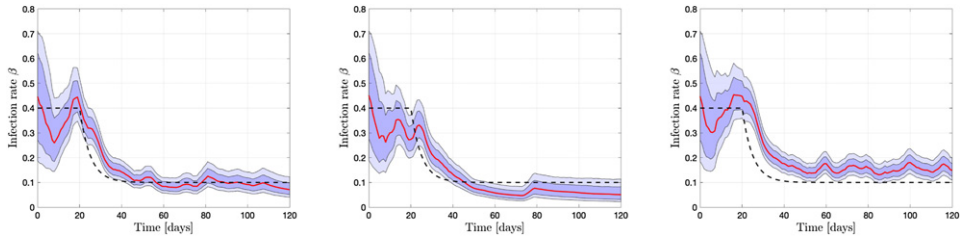

Figure 3. The particle estimates of the infection rate β corresponding to protocol A (left), B (middle) and C (right). The true generative model is plotted as a dashed curve, and the envelopes indicate the 50% and 75% credible intervals, the red curve being the median.

and without the scaling by $N_{\rm p}$, the innovation in a large population could be significantly large. Numerical tests indicate that with too large of an innovation, the algorithm may go astray.

The bar plots in figure 2 show the simulated numbers of new daily infections that are, from left to right, the data for protocols A, B and C. As expected, when the values of the static parameters increase, i.e., when the rate at which exposed/asymptomatic individuals become symptomatic or recover, is higher, the number of new daily infections is smaller, as shown by the bar plot in the middle panel (protocol B). When the rates decrease, hence individuals remain infectious longer, the number of new daily infections is higher, as shown in the right panel (protocol C). The red curves are the median values of the predicted mean of the data $\varphi_2 = \eta E$, with the shadowed regions marking the 50% (darker) and 75% (lighter) credibility envelopes. In all three cases, the predicted state estimation is in very good agreement with the data, regardless of whether or not the values of the static parameters were set to the values used in the generative model.

The panels in figure 3 show the β used in the generative model, formula (17) (dashed curve), together with the median, and 50% and 75% credible intervals computed from the particles corresponding to the data shown in figure 2. We observe that at the beginning of the simulation when the case count is low, and in protocol B in particular, the estimate of β varies a lot, while otherwise the estimated parameter follow well the generative model. When the static parameters in the solver were fixed to smaller values than in the generative model, the estimate of β is also smaller than the true value, as can be seen in the middle panel (protocol B). As expected, when the values of the static parameters used in the solver are larger than those in

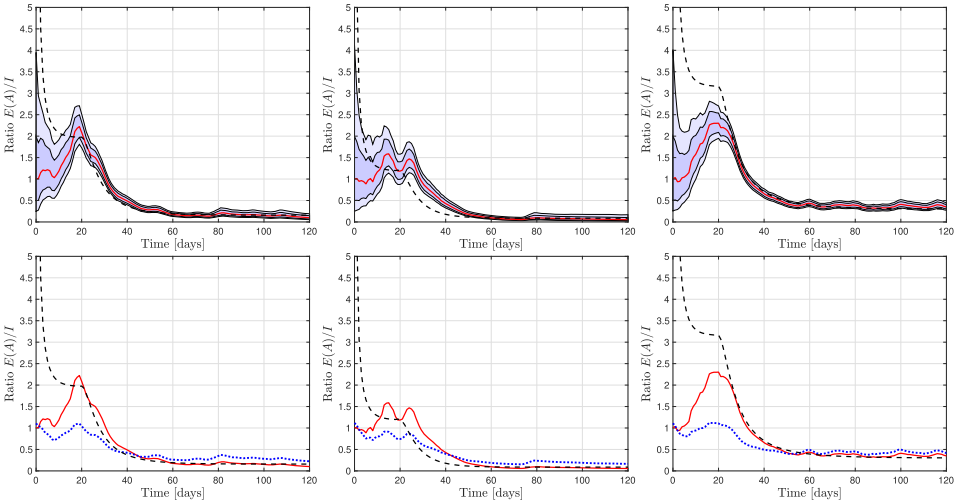

Figure 4. Upper row indicates the estimated ratio $\rho = E/I$ of asymtomatic and symptomatic cases in protocols A (left), B (middle) and C (right). The dashed curve is the true value of the generative model that the data were based on, and the red curve is the estimated median, enveloped by the 50% and 75% credible envelopes. In the lower row, the red and dashed curves are as in the upper row, while the blue dotted curve represents the equilibrium value ρ^* computed from the estimated median value of β and ν_S.

the generative model, the PF compensates by estimating a larger β: in the right panel (protocol C), the generative β is consistently fairly below the credible intervals.

The top row in figure 4 shows the evolution of the ratio ρ between the cohorts E and I as estimated from the state vectors via the PF with static γ and η parameters set, from left to right, equal to, smaller or larger than the values used in the generative model. As in the previous figure, the red curve shows the median, the shaded areas indicate the credibility intervals, and the dashed curve is the true ratio that the data simulations are based on. In all three cases, after the initial oscillations corresponding to few data, the behavior of the generative model is captured by the PF estimation with a delay of a few days, approximately matching the time needed to see a decrease of new daily infections. The bottom row shows, in addition to the median and true curves, also the equilibrium value ρ^*, formula (16), evaluated by using the estimated median values of β and ν_S. The plot indicates that at least in the waning phase of the infection, the ratio settles at or near the equilibrium value.

One of the most popular predictors of the course of an epidemic is the number of secondary infections per infectious capita. The time courses of the credible envelopes of the reproduction numbers $R_t^{\rm NG}$, R_t^I and R_t^E defined in section 3.3 for the three protocols are displayed in figure 5. In each panel, the dashed line is the R_t computed from the generative model, while the red curve and shaded regions are the median and credible intervals based on the PF estimates. In all cases, the estimate of R_t at the beginning of the epidemic is not stabilizing to the generative value, as could be expected because of the delay in the manifestation of the severity of the contagion in terms of individuals in the I cohort, and the low case count that is strongly affected by the Poisson noise. The same delay argument explains the shift by a few days of the median estimate with respect to the generative values, most noticeable at around day 20, when the number of secondary infections starts decreasing rather sharply. Indeed, for all three reproduction numbers, when the correct values of the static parameters are used, the

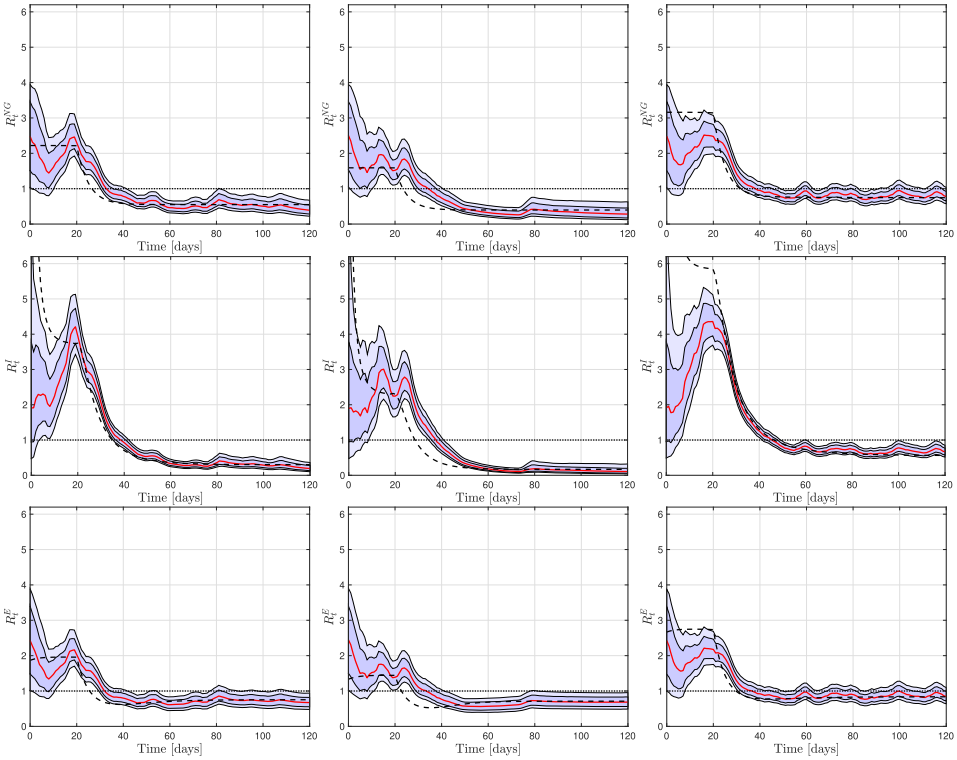

Figure 5. The different non-dimensional reproduction numbers estimated from the data in protocol A (left), B (middle) and C (right). From top to bottom, the R_t^{NG} based on the next generation matrix approach, the R_t^I controlling the local growth exponent of the symptomatic compartment I, and R_t^E controlling that of the asymptomatic cohort. As before, the median, the 50% and 75% credible envelopes are shown as well as the true generative value as a dashed curve.

generative curve is well within the credibility region, modulo a slight delay after peak incidence, as shown in the panels in the first column. The effect of using incorrect values of the static parameters can be seen during the uptick phase, when a putative longer time in the infectious pool leads to an overestimate of the secondary infections (second column) and a putative shorter time in the infectious pool leads to an overestimate (third column). Interestingly, the PF estimates capture quite accurately all three types of reproduction numbers in the waning phase of the epidemic, when the per capita number of secondary infections is less than one, regardless of whether the correct or incorrect values of the static parameters are used, corroborating that dimensionless quantities better compensate for modeling errors than dimensional ones such as β. A day-by-day comparison of the reproduction numbers for the same data set suggests that R_t^I is the most sensitive of the three to changes in the infectivity rate, with values up to 50% higher than the other two in the uptick phase, and smaller in the waning phase. The reproduction number R_t^{NG}, calculated on the basis of the next generation matrix, also captures well the trend of the pandemic, its decreasing following, with the delay of a few days, the slowing down of the infection rate, while the range of numerical values of R_t^E is the narrowest.

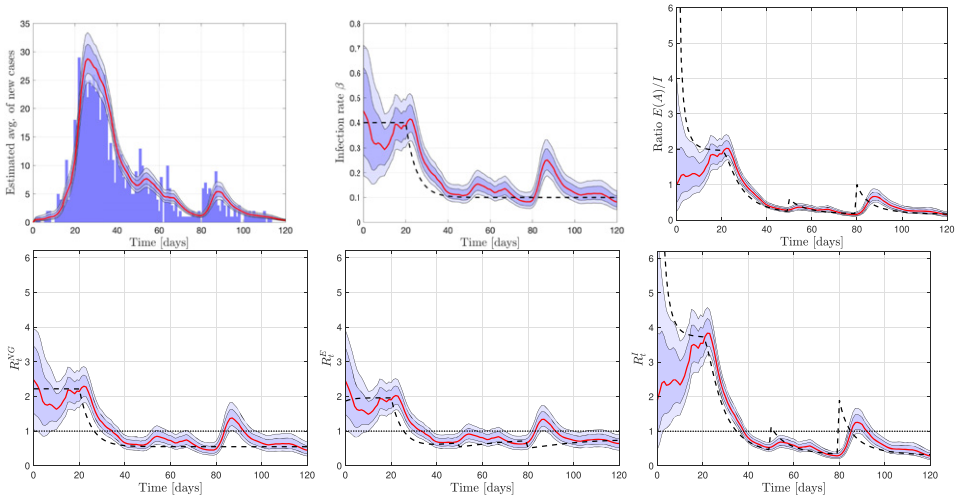

Figure 6. Effect of immigration: at times $t_1 = 50$ and $t_2 = 80$, 300 and 280 exposed (asymptomatic) individuals are imported from outside the system directly into the E compartment.

Example 2. The propagation model of the PF algorithm assumes implicitly a closed population, in which no exchange of individuals with the surrounding world takes place, a hardly tenable position in reality. The aim of the second example is to see, with simulated data, the effect on the state and parameter estimation of external influxes in the E pool, tantamount to importing a number of infectious asymptomatic individuals from a different, not tracked, pool. To isolate the effect, we assume that γ and η are known, and in our PF we set their values equal to those used for the generative model (protocol A). At time $t_1 = 50$ d, 300 asymptomatic individuals are imported into the E pool from outside the system, thus supplementing the flux coming from the susceptible compartment, to simulate the effect of traveling across different geographical regions, and the phenomenon is repeated at time $t_2 = 80$ d, when another bout of 280 asymptomatic cases is imported from outside the system. The PF for estimating the state and β parameter is not modified to account for the two bouts of immigration, and it receives the information only indirectly through the data. The two injections of asymptomatic cases are reflected, with a delay proportional to η^{-1}, in the data in the form of a bump in new daily symptomatic infections. A comparison of the top left panel of figure 6 and the left panel of figure 2 shows two bumps in the former shortly after $t = 50$ and $t = 80$. Although the immigration size is larger at $t = 50$, the effect of immigration at time $t = 80$ is more clearly visible, because of the lower prevalence: higher infection rate is needed to explain the growth with so few infectious individuals in the population. On the other hand, the effects of the immigration at time $t = 50$ last longer than those of the immigration at time $t = 80$, highlighting how the number of new cases in the days to come depends on both the transmission rate and the infectious pool size. The slight increase in the estimated β at time $t = 55$ and the more marked increase at time $t = 85$ show that, in the lack of better explanation, the PF attributes the unexpected change in the size of the I compartment to an increased transmission rate, as seen in the second panel of the first row of the same figure. This is in line with the understanding that the parameter models both the characteristic of the pathogen and the behavior of the underlying population.

Table 2. The parameter values used in the particle filter algorithm with unknown and not fixed static parameters.

Particle filter parameters		
a priori lower bound transmission rate [1/days]	$\beta_{\min}$	0.1
a priori upper bound for transmission rate	$\beta_{\max}$	0.6
a priori lower bound recovery rate [1/days]	$\gamma_{\min}$	1/7
a priori upper bound for recovery rate	$\gamma_{\max}$	1/28
a priori lower bound for incubation rate [1/days]	$\eta_{\min}$	1/15
a priori upper bound for incubation rate	$\eta_{\max}$	1/3
a priori upper bound of initial infectious individuals	$\lambda_{\max}$	5
Number of particles	N	20 000
Standard deviation of the propagation model for β	α	0.1
Standard deviation of innovation of the state	σ_1	0.01
Standard deviation of innovation of the transmission rate	σ_2	0.1

In addition, part of the unexplained increase in new cases is absorbed by the innovation of the state vector.

Changes in the trend of the epidemic are captured well by the estimates of all three reproduction numbers, shown in the bottom row of figure 6, which show a slight bump a few days after $t = 50$, and a more marked one a few days after $t = 80$. It is interesting that following the second immigration event, the reproduction numbers are estimated to exceed one, thus temporarily signaling an uptick in the number of infections. Since in this case the change in trend was due to an external event rather than a change in the population behavior, after a few days the R_t returns below one, while if the change in the data had been caused by the population becoming more complacent, this would have been the start of a new flare up.

Example 3. The third example illustrates how the particle filtering algorithm can estimate both the dynamic parameter β and the static parameters γ and η. The synthetic data (not shown) are generated using the generative model of example 1, protocol A. For the particle filtering algorithm with unknown static parameters, we set the parameters of the algorithm as in table 2.

The panels in the top row of figure 7 show the PF estimates with median and credibility envelopes, of the parameters β, γ and η together with the values used in the generative models. The estimate of β is rather close to the estimates found with fixed and correct static parameters, with a slightly increased posterior variance, and the estimate of γ improves as more data become available. The estimated reproduction numbers, displayed in the bottom row, capture again quite accurately the trend of the epidemic. The few days delay is due to the fact that the data used in the PF consist only of the reported infections, which typically follow by η^{-1} days' delay.

4.2. Testing with real data

In this section, the algorithm is tested with real COVID-19 infection data, downloaded from the website USAfacts (https://usafacts.org/) giving the daily COVID-19 cases in all US counties. In the first test, we consider the cases in Meade County, South Dakota, a relatively small community, whose population, according to the cited site, is $N_{\rm p} = 28\,332$. The county is the

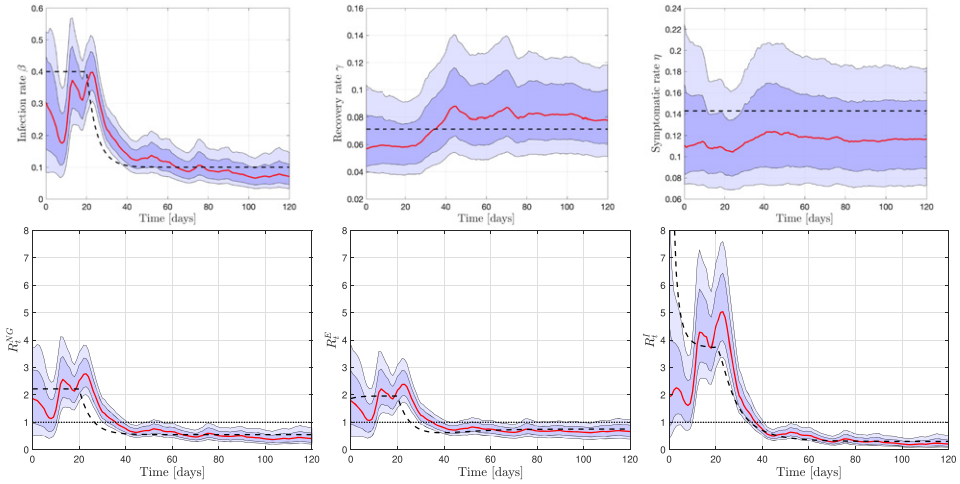

Figure 7. All three parameters β, γ and η are estimated. The top row shows the particle estimates of β, γ and η. The bottom row shows the particle estimates of the three reproduction numbers. In each plot, the dashed line indicates the values corresponding to the generative model.

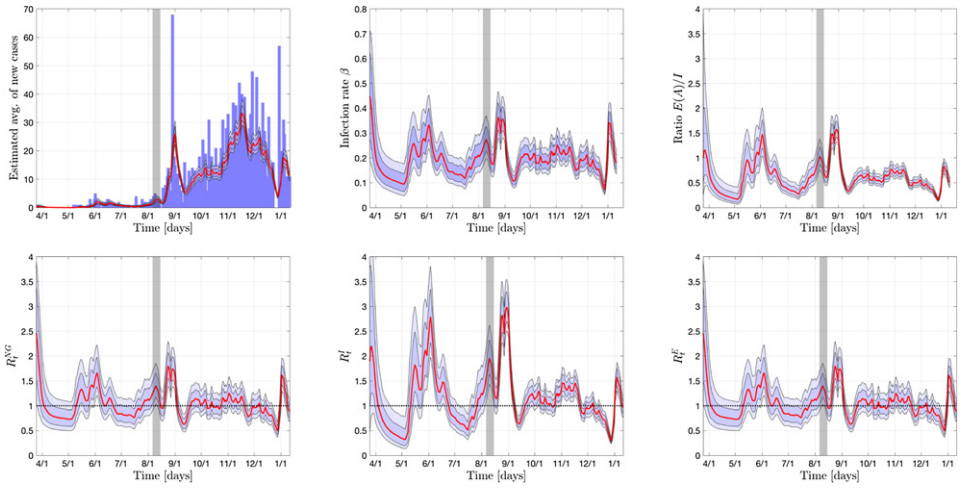

Figure 8. Particle filter estimates based on the COVID-19 data from Meade County, South Dakota. The gray shading marks the week of the Sturgis Motorbike Rally 2020 that brought approximately half a million bikers from all over the nation to Sturgis, the largest town in the county.

home of the town of Sturgis, famous for the annual motorbike reunion. The Sturgis Motorbike Rally 2020 was organized August 7th through 16th, drawing nearly half a million bikers from all over the nation, and it was predicted to be a potential superspreader event. As such, it provides a good example of the effect an external influx of asymptomatic infected individuals into a community with otherwise low prevalence, effectively becoming a real world version

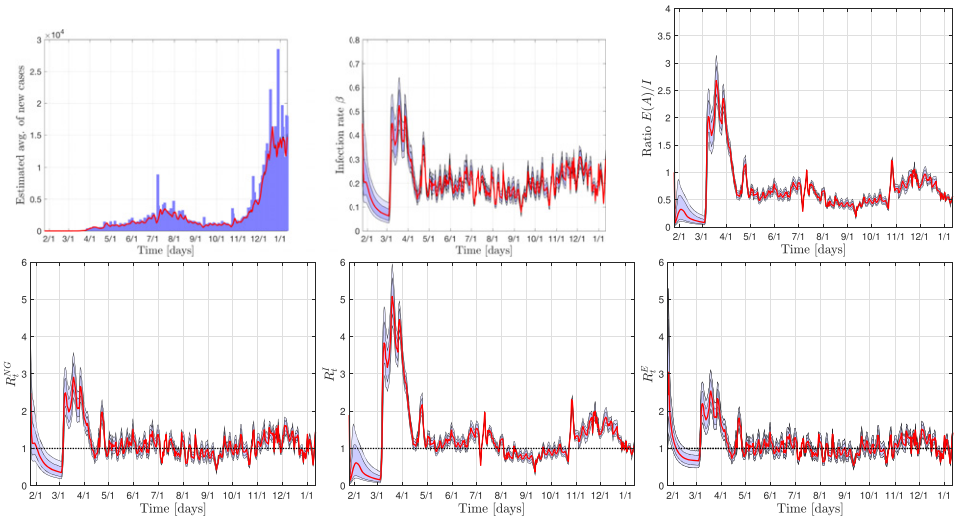

Figure 9. Particle filter estimates based on the COVID-19 data from Los Angeles County, California.

of example 2 of the previous section. We run the particle filtering algorithm keeping the static parameters fixed and equal to the values used in example 3.

The results show a significant increase in new daily infections roughly 1–2 weeks after the event, and the effect on β as well as the different R_t-numbers is similar to what was observed with the simulated data. Observe also that after the Motorbike Rally event, while the estimated β is not significantly larger than before, the case count did not return to the levels before the event, arguably suggesting that the event had a much larger negative impact on the community than just the immediate increase in cases. The data show also some reporting anomalies toward the end of December and early January. There is a significant uptick of cases after the reporting gap, and it is not clear whether the increase in estimated β and R_ts is purely a reporting artifact or an effect of gatherings during the holidays (figure 8).

In the last computed example, we apply the PF algorithm to data coming from Los Angeles County, California, with population over 10 million. The only parameter that was changed is the maximum initial number of infected and asymptomatic individuals, to scale it to the much larger population: we set $\lambda = 500$, reflecting the fact that the population is two orders of magnitude larger than in the previous example. As in the previous example, the first infected case was an isolated event (late January 2020), while the significant onset of the epidemic occurred in early March. The overall pattern of the estimates is similar to what was observed for the smaller community, however, we see that the estimates show a significant oscillatory pattern, reflecting the large deviations in the input data (see figure 9). This phenomenon in turn reflects the large population size, and signals the presence of different noise factors not taken into account in the propagation model that at least locally assumes a well mixed population. In a large population, reporting uncertainties become significant, while our model only assumes Poisson distributed noise.

We close this section with some observations on the model. One of the most important quantities for addressing the current COVID-19 pandemic is the ratio ρ of asymptomatic and symptomatic individuals. With real data, the particle filtering algorithm yields an estimate of ρ that is close to the value suggest by the C.D.C. at the time of writing this article, i.e., somewhere

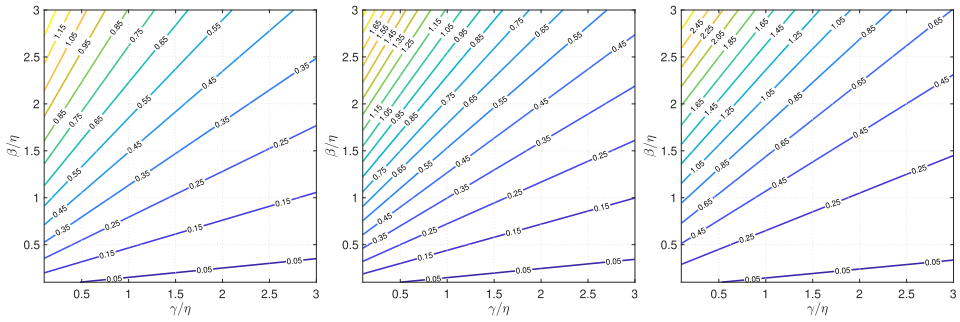

Figure 10. The equilibrium value ρ^* of the ratio E/I as a function of the dimensionless quantities $t = \gamma/\eta$ and $s = \beta/\eta$ with three different values of the infectivity factor of the cohort I: $q = 0.1$ (left), $q = 0.5$ (center) and $q = 1$ (right).

near 0.5. A natural question is whether this estimate depends significantly on the somewhat ad hoc parameter $q = 0.1$ used in the transmission model to weigh the contribution of symptomatic individuals, compared to the coefficient $p = 1$ used for the asymptomatic ones. To answer at least partially that question, consider the equilibrium value (16). To better understand the dependency of the ratio ρ^* on the model parameters, we introduce two dimensionless quantities characterizing the system of differential equations,

$$t = \frac{\gamma}{\eta}, \qquad s = \frac{\beta}{\eta}.$$

Neglecting the effect of the death rate μ, and assuming that the infection is not yet widespread, so that $S/N_{\rm p} \approx 1$, we find an approximate formula for ρ^* of (16) in terms of the dimensionless quantities t and s,

$$\rho^* = -\frac{1}{2}(1 + 2t - sq) + \sqrt{\frac{1}{4}(1 + 2t - sq)^2 + sp}, \tag{18}$$

Figure 10 shows the equilibrium value as a function of the dimensionless parameters for the three different choices $q = 0.1$ (left), $q = 0.5$ (center) and $q = 1$ (right), assuming that $p = 1$, indicating a relatively weak dependency of ρ^* on q. The quantity increases as s increases (higher infection rate) and decreases as γ increases (higher recovery rate). By using the values $\gamma = 1/14$ and $\eta = 1/7$, we have $t = 0.5$, implying that, neglecting the death rate, the equilibrium value $\rho^* = 0.5$ would correspond to effective equilibrium reproduction number

$$R_t^I \approx \frac{\rho^*}{t} = 1,$$

indicating that the symptomatic pool is near an equilibrium.

5. Discussion

The proposed dynamical Bayesian filtering algorithm is shown to provide a robust and consistent estimate of the state vectors, comprising the four cohorts of a new SE(A)IR model for COVID-19 spread, the transmission rate, a key parameter that is allowed to vary over time, and optionally the static parameters of recovery rate and the rate of symptom development. In epidemiology models, the transmission rate β is usually defined as $\beta =$ (transmission probability

per contact) $\times$ (number of contacts per day), where the first factor reflects the characteristics of the pathogen and is subject to change if new variants of the virus emerge, while the second factor is related to the behavior of the individuals and can change in connection with hygiene, social distancing and other mitigation measures. These factors justify the modeling of β as a time dependent parameter, and monitoring β in the light of infection data can be used as a measure of success of different mitigation measures, as well as a signal of new mutations taking ground. As shown by the computed examples, the parameter β is sensitive to mobility of the population, as the model considered in this article assumes a closed population. In [42], the current model was extended to a metapopulation model connecting different county level communities into an infection network.

Our SE(A)IR model directly addresses the role of virus shedding in transmission dynamics by asymptomatic individuals, many of whom recover before developing symptoms. We also provide a way for dynamically estimating the ratio of asymptomatic/presymptomatic/oligosymptomatic individuals in the E compartment to symptomatic individuals in the I compartment. Over a wide range of tests (see, e.g., [43] and the prediction website https://case.edu/medicine/healthintegration/covid-19-models) with real data, this ratio seems to settle consistently at values around or less than one, which is in accordance with the current understanding of the COVID-19 dynamics. The proposed SE(A)IR model does not address some of the features of COVID-19 transmission, including the latent period of the asymptomatic cohort. While designing a model, such as the SEAIR model (see, e.g., [17]) that properly accounts for the delay of the infectious phase of the asymptomatic patients is straightforward, retaining a structure that allows us to inform about the size of asymptomatic cohort based on data on symptomatic patients alone may be more challenging. Elaborating on that point will be a future direction of this research.

Acknowledgments

The work of DC was partly supported by NSF Grants DMS-1522334 and DMS-1951446, and the work of ES was partly supported by NSF Grant DMS-1714617.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://usafacts.org.

Appendix A

In this appendix we present a step by step derivation, explanation and interpretation of the particle filtering algorithms. First, consider a probability density $\pi(x)$ in $\mathbb{R}^n$, assume that $X^j \sim \pi$ are independent and identically distributed random variables, $j = 1, 2, \ldots$, and let φ be a continuous bounded function. By the strong law of large numbers, we know that as $N \to \infty$, almost surely,

$$\frac{1}{N}\sum_{j=1}^{N} \varphi(X^j) \to \int \varphi(x)\pi(x)\mathrm{d}x.$$

In particular, if x^j are realizations of X^j, i.e., independent random draws from π, we have that

$$\frac{1}{N}\sum_{j=1}^{N}\varphi(x^j) = \int \underbrace{\left(\sum_{j=1}^{N} w^j \delta_{x^j}(x)\right)}_{=\pi^N} \varphi(x)\mathrm{d}x \to \int \varphi(x)\pi(x)\mathrm{d}x \text{ almost surely,}$$

where $\delta_{x^j}(x)$ is a point measure at x^j and the weights w^j are uniform, $w^j = 1/N$. Hence, the discrete measures π^N converges weakly to π almost surely, which gives the precise meaning for the discrete approximation.

Consider now two probability densities π_1 and π_2 such that

$$\pi_1(x) = Cf(x)\pi_2(x), \tag{19}$$

where $f(x) \geqslant 0$ is measurable, and $C > 0$ is a, possibly unknown, normalizing constant. Let $\{x^1, x^2, \ldots, x^N\}$ be an independent sample (in the sense of the discussion above) from the distribution π_2, and let φ be a bounded continuous function. Using the weak approximation discussed above, we may write

$$\begin{aligned}\int \varphi(x)\pi_1(x)\mathrm{d}x = C\int \varphi(x)f(x)\pi_2(x)\mathrm{d}x &\approx \sum_{j=1}^{N} \underbrace{\frac{C}{N}f(x^j)}_{=w^j}\varphi(x^j)\\ &= \sum_{j=1}^{N} w^j\varphi(x^j),\end{aligned} \tag{20}$$

that is, we have an approximation

$$\pi_1(x)\mathrm{d}x \approx \sum_{j=1}^{N} w^j\delta_{x^j}(x), \quad w^j = \frac{C}{N}f(x^j).$$

Observe that, if C is unknown, so are the weights. However, since π_1 is a probability measure, by substituting $\varphi(x) = 1$ in (20) implies that the weights can be calculated as

$$w^j = \frac{\widehat{w}^j}{\sum \widehat{w}^\ell}, \qquad \widehat{w}^j = f(x^j).$$

This forms the foundation of a practical scheme for drawing from a distribution π_1 of form (19), provided that we have a method to draw independent samples from π_2.

Algorithm 1:

(a) Draw an independent sample from π_2; denote the sample points by $x^1, x^2, \ldots, x^N$.
(b) Calculate the weights $w^j = f(x^j)$.
(c) Normalize, setting

$$w^j \to \frac{w^j}{\sum w^j}.$$

To put this in the context of our PF algorithm, we need to add one more layer. Consider a probability density π that can expressed as

$$\pi(x) = \sum_{j=1}^{N} g_j \pi_j(x), \quad g_j \geqslant 0, \tag{21}$$

where the π_j are probability densities. Integrating over the whole space and recalling that a probability density is normalized so that its integral is one,

$$\sum_{j=1}^{N} g_j = 1,$$

that is, π is a convex combination of the π_j. Next we interpret the right-hand side of (21) as a marginal probability. In fact, if J denote a discrete random variable with point mass $P(J = j) = g_j$, and $\pi_j(x) = \pi_{X|J}(x|j)$, then

$$\pi(x) = \sum_{j=1}^{N} P(J = j)\pi_{X|J}(x|j).$$

In light of the above, we arrive at the following two-phase scheme for drawing from the density (21).

Algorithm 2:

(a) Draw an index j, $0 \leqslant j \leqslant$ using the coefficients g_j as point mass probabilities;
(b) Given j, draw x from π_j according to algorithm 1.

Combing algorithms 1 and 2, we arrive at the particle filtering algorithm described in the article. Indeed,

$$\pi_{X_{t+1}|\mathscr{B}_{t+1}}(x_{t+1}|\mathcal{B}_{t+1}) \approx \sum_{j=1}^{N} \underbrace{\left\{ w_t^j \pi_{B_{t+1}|X_{t+1}}(b_{t+1}|\widehat{x}_{t+1}^j) \right\}}_{\text{(a)}} \times \underbrace{\left\{ \frac{\pi_{B_{t+1}|X_{t+1}}(b_{t+1}|x_{t+1})}{\pi_{B_{t+1}|X_{t+1}}(b_{t+1}|\widehat{x}_{t+1}^j)} \right\}}_{\text{(b)}} \underbrace{\pi_{X_{t+1}|X_t}(x_{t+1}|x_t^j)}_{\text{(c)}},$$

where (a) plays the role of g_j. Once we have drawn j according to the first step of algorithm 2, we are ready to draw from the density

$$\pi_j(x_{t+1}) = \underbrace{\left\{ \frac{\pi_{B_{t+1}|X_{t+1}}(b_{t+1}|x_{t+1})}{\pi_{B_{t+1}|X_{t+1}}(b_{t+1}|\widehat{x}_{t+1}^j)} \right\}}_{\text{(b)}} \underbrace{\pi_{X_{t+1}|X_t}(x_{t+1}|x_t^j)}_{\text{(c)}}.$$

Here, (b) plays the role of $Cf(x)$ in (19), and (c) plays the role π_2 in algorithm 1.

Appendix B

In this appendix we discuss in detail selection of the innovation model for the static parameters. For simplicity, we denote $\xi = \log\,\theta \in \mathbb{R}^2$. Assume that we use an innovation model similar to

X_t, that is,

$$\Xi_{t+1} = \xi_t + h\Sigma_{t+1}^{1/2} W_{t+1}, \quad W_{t+1} \sim \mathcal{N}(0, \mathrm{I}_2), \tag{22}$$

or,

$$\pi_{\Xi_{t+1}|\Xi_t}(\xi_{t+1}|\xi_t) = \mathcal{N}(\xi_{t+1}|\xi_t, h^2\Sigma_{t+1}),$$

where Σ_{t+1} is the empirical covariance calculated by using the predictive particles $\widehat{\xi}_{t+1}^j$. We extend the stochastic model to comprise the parameter ξ also, and write the approximation

$$\pi_{X_t,\Xi_t|\mathcal{B}_t}(x_t, \xi_t|\mathfrak{B}_t) \approx \sum_{j=1}^N w_t^j \delta_{x_t^j}(x_t)\delta_{\xi_t^j}(\xi_t).$$

Replacing the density with its particle approximation in the Chapman–Kolmogorov formula (8) yields

$$\begin{aligned}\pi_{X_{t+1},\Xi_{t+1}|\mathscr{B}_t}(x_{t+1}, \xi_{t+1}|\mathfrak{B}_t) &= \iint \pi_{X_{t+1}|X_t}(x_{t+1}|x_t)\pi_{\Xi_{t+1}|\Xi_t}(\xi_{t+1}|\xi_t)\pi_{X_t|\mathscr{B}_t}(x_t|\mathfrak{B}_t)\mathrm{d}x_t\mathrm{d}\xi_t \\ &\approx \sum_{j=1}^N w_t^j \pi_{X_{t+1}|X_t}(x_{t+1}|x_t^j)\mathcal{N}(\xi_{t+1}|\xi_t^j, h^2\Sigma_{t+1}).\end{aligned} \tag{23}$$

Marginalizing with respect to X_{t+1}, we obtain the approximation of the prior density of Ξ_{t+1},

$$\begin{aligned}\pi_{\Xi_{t+1}|\mathscr{B}_t}(\xi_{t+1}|\mathfrak{B}_t) &= \int \pi_{X_{t+1},\Xi_{t+1}|\mathscr{B}_t}(x_{t+1}, \xi_{t+1}|\mathfrak{B}_t)\mathrm{d}x_{t+1} \\ &\approx \sum_{j=1}^N w_t^j \mathcal{N}(\xi_{t+1}|\xi_t^j, h^2\Sigma_{t+1}),\end{aligned}$$

as a Gaussian mixture. The mean of the Gaussian mixture is

$$\overline{\xi}_{t+1} \approx \sum_{j=1}^N w_t^j \int \mathcal{N}(\xi_{t+1}|\xi_t^j, h^2\Sigma_{t+1})\mathrm{d}\xi_{t+1} = \sum_{j=1}^N w_t^j \xi_t^j = \overline{\xi}_t,$$

and it can be verified that the covariance is

$$\begin{aligned}\mathrm{Cov}(\Xi_{t+1}) &\approx \sum_{j=1}^N w_j \int (\xi_{t+1} - \overline{\xi}_t)(\xi_{t+1} - \overline{\xi}_t)^{\mathsf{T}} \mathcal{N}(\xi_{t+1}|\xi_t^j, h^2\Sigma_{t+1}) \\ &= h^2\Sigma_{t+1} + \sum_{j=1}^N w_j(\xi_t^j - \overline{\xi}_t)(\xi_t^j - \overline{\xi}_t)^{\mathsf{T}} = (1 + h^2)\Sigma_{t+1}.\end{aligned}$$

Therefore the innovation model (22) increases the covariance of parameter by a factor of $1 + h^2$, in conflict with the prior assumption that these parameters are static. One way to compensate for this artificial diffusion, as suggested in [35, 44], is to shift slightly the centers of the Gaussians in the mixture towards the ensemble mean, that is,

$$\pi_{\Xi_{t+1}|\Xi_t}(\xi_{t+1}|\xi_t) = \mathcal{N}(\xi_{t+1}|a\xi_t + (1-a)\overline{\xi}_t, h^2\Sigma_{t+1}), \quad a^2 + h^2 = 1.$$

It is straightforward to verify that this innovation model preserves the particle-based mean and covariance.

ORCID iDs

D Calvetti https://orcid.org/0000-0001-5696-718X
A Hoover https://orcid.org/0000-0001-8812-0354
J Rose https://orcid.org/0000-0001-7458-9303
E Somersalo https://orcid.org/0000-0001-5099-3512

References

[1] Ogilvy Kermack W and McKendrick A G 1927 A contribution to the mathematical theory of epidemics *Proc. R. Soc.* A **115** 700 –21 Containing papers of a mathematical and physical character
[2] Bendavid E *et al* 2021 COVID-19 antibody seroprevalence in Santa Clara County, California *Int. J. Epidemiol.* **1** 10
[3] Streeck H, Hartmann G, Exner M and Schmid M 2020 Preliminary result and conclusions of the COVID-19 case cluster study (Gangelt Municipality) See https://land.nrw/sites/default/files/asset/document/zwischenergebnis_covid19_case_study_gangelt_en.pdf
[4] Bohk-Ewald C, Dudel C and Myrskylä M 2020 A demographic scaling model for estimating the total number of COVID-19 infections (arXiv:2004.12836)
[5] MacDonald R G 1952 The analysis of equilibrium in malaria *Tropical Dis. Bull.*
[6] Diekmann O, Heesterbeek J A P and Metz J A J 1990 On the definition and the computation of the basic reproduction ratio r_0 in models for infectious diseases in heterogeneous populations *J. Math. Biol.* **28** 365–82
[7] Diekmann O, Heesterbeek J A P and Roberts M G 2010 The construction of next-generation matrices for compartmental epidemic models *J. R. Soc. Interface.* **7** 873–85
[8] Heesterbeek J A P 2002 A brief history of R_0 and a recipe for its calculation *Acta Biotheoretica* **50** 189–204
[9] Heffernan J M, Smith R J and Wahl L M 2005 Perspectives on the basic reproductive ratio *J. R. Soc. Interface.* **2** 281–93
[10] Dietz K 1993 The estimation of the basic reproduction number for infectious diseases *Stat. Methods Med. Res.* **2** 23–41
[11] Brauer F, Castillo-Chavez C and Feng Z 2019 *Mathematical Models in Epidemiology* (Berlin: Springer)
[12] Park M, Cook A R, Lim J T, Sun Y and Dickens B L 2020 A systematic review of COVID-19 epidemiology based on current evidence *J. Clin. Med.* **9** 967
[13] Edelstein-Keshet L 2005 *Mathematical Models in Biology* (Philadelphia, PA: SIAM)
[14] De Vries G, Hillen T, Lewis M, Müller J and Schönfisch B 2006 *A Course in Mathematical Biology: Quantitative Modeling with Mathematical and Computational Methods* (Philadelphia, PA: SIAM)
[15] Calvetti D and Somersalo E 2012 *Computational Mathematical Modeling: An Integrated Approach across Scales* vol 17 (Philadelphia, PA: SIAM)
[16] Murray C J L *et al* (IHME COVID) 2020 Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months MedRxiv
[17] Liu Z, Magal P, Seydi O and Webb G 2020 Predicting the cumulative number of cases for the COVID-19 epidemic in China from early data (arXiv:2002.12298)
[18] He S, Peng Y and Sun K 2020 SEIR modeling of the COVID-19 and its dynamics *Nonlinear Dyn.* **101** 1667–80
[19] Marangoni A G 2003 *Enzyme Kinetics: A Modern Approach* (New York: Wiley)
[20] Brynjarsdóttir J and O'Hagan A 2014 Learning about physical parameters: the importance of model discrepancy *Inverse Problems* **30** 114007

[21] Comunian A, Gaburro R and Giudici M 2020 Inversion of a SIR-based model: a critical analysis about the application to COVID-19 epidemic *Physica* D **413** 132674
[22] Bärwolff G 2020 Prospects and limits of SIR-type mathematical models to capture the COVID-19 pandemic (arXiv:2004.06522)
[23] Prodanov D 2021 Analytical parameter estimation of the SIR epidemic model. Applications to the COVID-19 pandemic *Entropy* **23** 59
[24] Senel K, Ozdinc M and Ozturkcan S 2020 Single parameter estimation approach for robust estimation of SIR model with limited and noisy data: the case for COVID-19 *Disaster Med. Public Health Preparedness* 1–15
[25] Pacheco C C and de Lacerda C R 2021 Function estimation and regularization in the SIRD model applied to the COVID-19 pandemics *Inverse Problems Sci. Eng.* 1–16
[26] Marinov T T and Marinova R S 2020 Dynamics of COVID-19 using inverse problem for coefficient identification in SIR epidemic models *Chaos Solitons Fractals* X **5** 100041
[27] Hong H G and Li Y 2020 Estimation of time-varying reproduction numbers underlying epidemiological processes: a new statistical tool for the COVID-19 pandemic *PloS One* **15** e0236464
[28] Kabanikhin S I and Krivorotko O I 2020 Mathematical modeling of the Wuhan COVID-2019 epidemic and inverse problems *Comput. Math. Math. Phys.* **60** 1889–99
[29] Kucharski A J *et al* 2020 Early dynamics of transmission and control of COVID-19: a mathematical modelling study *Lancet Infectious Diseases* **20** 553
[30] Yang H and Lee J 2020 Variational Bayes method for ODE parameter estimation with application to time-varying SIR model for COVID-19 epidemic (arXiv:2011.09718)
[31] Hao X, Cheng S, Wu D, Wu T, Lin X and Wang C 2020 Reconstruction of the full transmission dynamics of COVID-19 in Wuhan *Nature* **584** 420–4
[32] Engbert R, Rabe M M, Kliegl R and Reich S 2021 Sequential data assimilation of the stochastic SEIR epidemic model for regional COVID-19 dynamics *Bull. Math. Biol.* **83** 1–16
[33] Leung K, Wu J T, Liu D and Leung G M 2020 First-wave COVID-19 transmissibility and severity in China outside Hubei after control measures, and second-wave scenario planning: a modelling impact assessment *Lancet* **395** 1382
[34] Safta C, Ray J and Sargsyan K 2020 Characterization of partially observed epidemics through Bayesian inference: application to COVID-19 *Comput. Mech.* **66** 1109–29
[35] Liu J and West M 2001 Combined parameter and state estimation in simulation-based filtering *Sequential Monte Carlo Methods in Practice* (Berlin: Springer) pp 197–223
[36] Kaipio J and Somersalo E 2006 *Statistical and Computational Inverse Problems* vol 160 (Berlin: Springer)
[37] Doucet A and Johansen A M 2009 A tutorial on particle filtering and smoothing: fifteen years later *Handbook of Nonlinear Filtering* vol 12 p 3
[38] Arnold A, Calvetti D and Somersalo E 2014 Parameter estimation for stiff deterministic dynamical systems via ensemble Kalman filter *Inverse Problems* **30** 105008
[39] Arnold A, Calvetti D and Somersalo E 2013 Linear multistep methods, particle filtering and sequential Monte Carlo *Inverse Problems* **29** 085007
[40] Arnold A, Calvetti D, Gjedde A, Iversen P and Somersalo E 2015 Astrocytic tracer dynamics estimated from [1-11C]-acetate PET measurements *Math. Med. Biol. J. IMA* **32** 367–82
[41] Fearnhead P and Künsch H R 2018 Particle filters and data assimilation *Annu. Rev. Stat. Appl.* **5** 421–49
[42] Calvetti D, Hoover A P, Rose J and Somersalo E 2020 Metapopulation network models for understanding, predicting, and managing the coronavirus disease COVID-19 *Front. Phys.* **8** 261
[43] Calvetti D, Hoover A, Rose J and Somersalo E 2020 Bayesian dynamical estimation of the parameters of an SE(A)IR COVID-19 spread model (arXiv:2005.04365)
[44] West M 1993 Approximating posterior distributions by mixtures *J. R. Stat. Soc.* B **55** 409–22