Inference for partially observed epidemic dynamics guided by Kalman filtering techniques

Romain Narci, Maud Delattre, Catherine Larédo, Elisabeta Vergu

Inference for partially observed epidemic dynamics guided by Kalman filtering techniques [†]

Romain Narci[a,*], Maud Delattre[a], Catherine Larédo[a], Elisabeta Vergu[a]

[a] *MaIAGE, INRAE, Université Paris-Saclay, 78350 Jouy-en-Josas, France*

Abstract

Despite the recent development of methods dealing with partially observed epidemic dynamics (unobserved model coordinates, discrete and noisy outbreak data), some limitations remain in practice, mainly related to the amount of augmented data and the calibration of numerous tuning parameters. In particular, coordinates of dynamic epidemic models being coupled, the presence of unobserved coordinates leads to a statistically difficult problem. A generic inference method easily practicable and able to tackle these issues is proposed. First, using the properties of epidemics in large populations, a two-layer model is built. Through a diffusion based approach, a Gaussian approximation of the epidemic density-dependent Markovian jump process is obtained, representing the state model. The observational model, consisting in noisy observations of some model coordinates, is approximated by Gaussian distributions. Then, an inference method based on an approximate likelihood using Kalman filter recursions is developed to estimate parameters of both state and observational models. Performances of estimators of key model parameters are assessed on simulated data of SIR epidemic dynamics for different scenarios with respect to the population size and the number of observations. These performances are compared with those obtained using the widely used method of Maximum Iterated Filtering (MIF). Finally, the inference method is applied on a real data set of influenza outbreak in a North England boarding school in 1978.

Keywords: Approximated maximum likelihood; Diffusion approach; Kalman filter; Measurement errors; Partially observed Markov process; Epidemic dynamics.

1. Introduction

The interest and impact of mathematical modeling and inference methods for infectious diseases have considerably grown in recent years in the context of increasing complexity of models and of abundant data with varying quality. Estimating the parameters governing epidemic dynamics from available data has become a major challenge, in particular from the perspective of subsequently providing reliable predictions of these dynamics. Many authors have addressed the problem of key epidemic parameters estimation based on likelihood approaches (e.g.

[†] R codes are provided as annexes.

[*] Corresponding author

Email address: romain.narci@inrae.fr (Romain Narci)

Cauchemez and Ferguson (2008)). While estimation is quite straighforward for complete observations, this is no longer true for incomplete observations set-up that occurs in practice, regardless of the mathematical formalism. Indeed, the available data are most often partially observed (e.g. some health statuses, such as asymptomatic infected stages, cannot be observed at all; infectious and recovery dates are not observed for all individuals during the outbreak, not all the infectious individuals are reported) and also temporally and/or spatially aggregated. Various approaches were developed to deal with such patterns of data (e.g. see O'Neill (2010), Britton and Giardina (2016) for reviews). In the general framework of the partially observed Markov processes, some of these methods have been implemented in the `R` package POMP (King et al. (2017)). Among these methods, we can mention the maximum iterated filtering (MIF: Ionides et al. (2006), Ionides et al. (2015)), for which the parameter space is explored by considering that parameters follow a random walk in time, with variance decreasing over the filtering iterations and the likelihood is stochastically estimated. Theoretical justification for convergence to the maximum likelihood estimates in the parameter space was provided for this method. (Ionides et al. (2011)). Besides, likelihood-free methods, such as the Approximate Bayesian Computation based on sequential Monte Carlo (ABC-SMC, Sisson et al. (2007), Toni et al. (2009)) or the Particle Markov chain Monte Carlo (PMCMC, Andrieu et al. (2010)), opened some of the most promising pathways for improvement. Nevertheless, these algorithms do not provide a definitive solution to the statistical inference from incomplete epidemic data. Indeed, there are some limitations in practice due to the amount of augmented data and to the adjustment of the numerous tuning parameters. That can lead to substantial computational effort.

In this paper, we propose a different approach to deal with the presence of missing coordinates, discrete observations, reporting and measurement errors.

Using the large population framework, *i.e.* where the population size is large, we first build an approximation of the epidemic dynamics represented by Markovian jump processes by an autoregressive Gaussian process through a diffusion approach (see e.g. Ethier and Kurtz (2005), Guy et al. (2015)). Then we account simultaneously for the lack of one coordinate and for the systematic noise present in observations by applying a projection operator to the process and by adding heteroscedastic Gaussian errors. This yields an approximate likelihood which contains the unobserved coordinate. We propose a method based on Kalman filtering in order to compute the approximated log-likelihood of the available observations and, consequently, to estimate model parameters. An first innovative aspect with respect to other inference methods are the use of Kalman filter to recursively compute the approximate likelihood in the non standard case of small noise framework, a setting involving a small parameter equal to the inverse of the square root of the population size, rather than the classical recurrent case coupled with large observation time-window (the number of observations going to infinity). Also, the explicit integration in the algorithm of the data sampling interval and an alternative view point in the prediction of the successive model states given the observations are two additional innovative points.

The derivation and accuracy assessment of Gaussian process approximations for stochastic epidemic models were already described in Buckingham-Jeffery et al. (2018) along with maximum likelihood inference for parameters underlying epidemic dynamics. However, this study does not rely on Kalman filtering nor considers noise in outbreak data. The computation of the approximate likelihood of the associated statistical model and parameter estimation, performed via Kalman filter recursions, was proposed in Favetto and Samson (2010), but for simpler models without nonlinear terms in the drift and with no parameter to estimate in the diffusion term.

For sake of simplicity, we consider here an epidemic with homogeneous mixing in a closed population, whose dynamics is described by a compartmental model, each compartment containing

individuals with identical health states. We focus on the simple SIR (susceptible - infectious - recovered) epidemic model, described by a two-dimensional jump process, partially observed at regularly spaced discrete times with measurement errors. The approach can be easily extended to broader epidemic models observed with various sampling intervals.
The paper is organized as follows. In Section 2 we introduce the general framework and the related inference issues, and propose the model approximation. Section 3 contains the main methodological improvements of our paper: the construction of the approximate log-likelihood, its computation based on Kalman filter recursions and the associated parameter estimation. In Sections 4 and 5, we assess the performances of our estimators, respectively on simulated data and on the real data of the influenza in a Britain boarding school in 1978, and compare our results with those obtained with the MIF method. Section 6 contains a discussion and concluding remarks.

2. Gaussian model approximations for large population epidemics

2.1. Preliminary comments on inference in epidemic models

Epidemic dynamics can be naturally described through compartmental models which are by essence mechanistic and include parameters in their description. In such models, the population is partitioned in compartments corresponding to different stages of the infection process whose temporal evolution is described. As an illustrative example all along the article, we will use the simple SIR epidemic model. At any time, each individual is either susceptible (S), infectious (I) or recovered (R). In this model, there are two mechanistic parameters of interest that govern the transitions of individuals between the states S, I and R: the transmission rate of the pathogen λ and the recovery rate γ. More precisely, individuals can move from state S to state I according to λ or move from state I to state R according to γ (Figure 1).

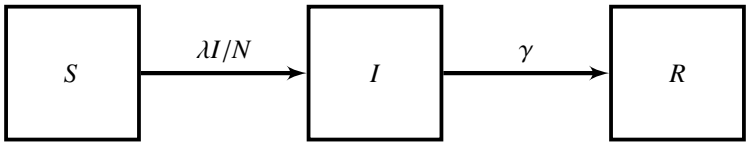

Figure 1: SIR compartmental model with three blocks corresponding respectively to susceptible (S), infectious (I) and recovered (R) individuals. Transitions of individuals from state S to state I are governed by the transmission rate λ and transitions of individuals from state I to state R are governed by the recovery rate γ of the epidemics.

One of the main objectives of epidemic studies is to estimate the related mechanistic parameters from the available data. One of the most natural probabilistic representations of compartmental epidemic models is the continuous-time Markov jump process (see Section 2.2). Inference for Markov jump processes is straightforward when the sample paths are completely observed. In the context of epidemics, this is equivalent with the observation of all infection and recovery times for all the individuals in the population. This rarely occurs in practice because very often, one or more coordinates (*i.e.* $S(t)$, $I(t)$) are not observed and because the available observations are only collected at discrete time points t_k with $0 = t_0 < t_1 < t_2 < \cdots < t_n = T$ on a finite time interval $[0, T]$. More specifically, the data often consist in the counting of the newly infected individuals $N_I(t_k)$ on successive time intervals $[t_{k-1}, t_k]$. Alternatively, the successive numbers of infectious individuals $I(t_k)$ are sometimes available, especially for low population sizes. Moreover, it is common that the available data are affected by several sources of noise, such as under-reporting

of infection events or, when reported, imperfect diagnostic tests. In a nutshell, the nature of the data makes it difficult to infer key epidemic parameters: (i) the observations are available at discrete time points, (ii) not all the coordinates of the dynamical model are observed and (iii) systematic reporting and measurement errors have to be taken into account. Our objective is to propose a convenient latent variable model formulation allowing for the estimation of key epidemic parameters from imperfect observations of outbreaks.

2.2. Approximation of large population epidemic models and an autoregressive view

Consider an epidemic in a closed population with homogeneous mixing modeled by a d-dimensional Markov jump process $\mathcal{Z}(t)$ where d is the number of compartments corresponding to the successive health status within the population. If N is the population size, the state space of $(\mathcal{Z}(t), t \geq 0)$ is $E = \{0, \dots, N\}^d$. Let $Q = (q_{k,l}, k, l \in E)$ denote its Q-matrix. It satisfies $\forall l \neq k,\ q_{k,l} \geq 0$, and $q_{k,k} = -\sum_{l \in E, l \neq k} q_{k,l}$. There are two standard ways of describing this jump process (see e.g. Norris (1997)):
(i) It is defined by the underlying jump chain and holding times. Starting from Q, set $\pi_{k,l} = \frac{q_{k,l}}{q_k}$ with $q_k = -q_{k,k}$, $\pi_{k,k} = 0$ if $q_k = 0$ and $\pi_{k,k} = 1$ if $q_k \neq 0$. The process stays in state k according to an exponential distribution $\mathcal{E}(q_k)$ and jumps to state l with probability $\pi_{k,l}$.
(ii) Its infinitesimal generator : as $h \to 0$, $\mathbb{P}(\mathcal{Z}(t+h) = l | \mathcal{Z}(t) = k) = \delta_{k,l} + q_{k,l}h + o_P(h)$, where $\delta_{k,l}$ denotes the Kronecker function ($\delta_{k,l} = 1$ if $l = k$, $\delta_{k,l} = 0$ if $l \neq k$). Hence, for f a measurable function $E \to \mathbb{R}$, if $\mathbb{E}_k$ denotes the expectation conditional on $\mathcal{Z}(0) = k$, $Qf(k) = \sum_{l \in E} q_{k,l} f(l) = \lim_{t \to 0} \frac{1}{t}(\mathbb{E}_k f(\mathcal{Z}(t)) - f(k))$. Simulations of $\mathcal{Z}(t)$ are usually based on (i), while (ii) relies on general properties of Markov processes.

For any vector V or any matrix M, let V^t and M^t denote the transpose vector or matrix. For $\ell \neq (0, \dots, 0)^t$ a jump of $\mathcal{Z}(t)$, define the jump function

$$\alpha_\ell(k) = q_{k,k+\ell} \quad \text{for } \ k, k+\ell \in E.$$

Consider now $(\mathcal{Z}_N(t))_{t \geq 0}$ the normalized Markov jump process,

$$\mathcal{Z}_N(t) = \frac{\mathcal{Z}(t)}{N} \in E^N = \{k/N, k \in E\}. \tag{1}$$

The associated jump functions are, for $x \in E^N$, $\alpha_\ell^N(x) = \frac{1}{N}\alpha_\ell([Nx])$. Assume that the process $(\mathcal{Z}(t))$ is density-dependent *i.e.*,

$$\begin{aligned} &\textbf{H1 :}\ \forall \ell,\ \ \forall x \in [0,1]^d,\ \frac{1}{N}\alpha_\ell([Nx]) \underset{N \to +\infty}{\to} \beta_\ell(x), \\ &\textbf{H2 :}\ \forall \ell,\ \ \beta_\ell \in C^2([0,1]^d, \mathbb{R}), \end{aligned}$$

where $[Nx]$ is the vector of integers $[Nx_1], ..., [Nx_d]$ with $[Nx_i]$ being the integer part of Nx_i. Define, for $x \in [0,1]^d$, the function $b(\cdot)$ and the $d \times d$ symmetric non negative matrix $\Sigma(\cdot)$:

$$b(x) = \sum_{\ell \in E^-} \ell\, \beta_\ell(x)\ ; \quad \Sigma(x) = \sum_{\ell \in E^-} \beta_\ell(x)\, \ell\, \ell^t. \tag{2}$$

For the SIR epidemic model in a closed population, we have that $S(t) + I(t) + R(t) = N$ for all t. Therefore, its state space is $E = \{0, \dots, N\}^2$. Only two jumps are possible from $k = (S, I)^t$:

- $\ell_1 = (-1, +1)^t$: $(S, I) \to (S-1, I+1) \Rightarrow q_{k,k+\ell_1} = \lambda S I/N = \alpha_{\ell_1}(k)$,
- $\ell_2 = (0, -1)^t$: $(S, I) \to (S, I-1) \Rightarrow q_{k,k+\ell_2} = \gamma I = \alpha_{\ell_2}(k)$.

This process is density dependent: if $s = \frac{S}{N}, i = \frac{I}{N}$, then, $\frac{1}{N}\alpha_{\ell_1}([Ns],[Ni]) = \frac{1}{N}(\lambda[Ns])\frac{[Ni]}{N} \to \lambda si$ and $\frac{1}{N}\alpha_{\ell_2}([Ns],[Ni]) = \frac{1}{N}\gamma[Ni] \to \gamma i$ as $N \to \infty$.
Moreover (2) is, for $x = \begin{pmatrix} s \\ i \end{pmatrix}$, $b(x) = \lambda si\begin{pmatrix} -1 \\ 1 \end{pmatrix} + \gamma i\begin{pmatrix} 0 \\ -1 \end{pmatrix}$, $\Sigma(x) = \lambda si\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} + \gamma i\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$.
Let us recall the Law of Large Numbers result stated for instance in Britton and Pardoux (2020), Part I, Theorem 2.2.7 page 31.

Lemma 1. *Assume that $(\mathcal{Z}(t))$ satisfies (H1), (H2) and $\mathcal{Z}_N(0) \to x_0$ as $N \to +\infty$. Then, $(\mathcal{Z}_N(t))$ converges almost surely uniformly on $[0, T]$ to the solution $x(t)$ of the ordinary differential equation*

$$\frac{dx}{dt} = b(x(t)); \quad x(0) = x_0. \tag{3}$$

If $x_0 = (0, \ldots, 0)^t$, then $x(t) = 0$ for all t and (3) does no longer describe adequately the epidemic dynamics (see e.g Britton and Pardoux (2020) Part I). Equation (3) describes the dynamics in case of a major outbreak corresponding to $x_0 \neq (0, \ldots, 0)^t$.
In Guy et al. (2015), by extending results of Ethier and Kurtz (2005), another approximation of the epidemic model is proposed leading to a diffusion process $(Z_N(t))_{t\geq 0}$ with small diffusion matrix $\frac{1}{\sqrt{N}}\Sigma(x)$ where Σ is the matrix defined in (2),

$$\begin{cases} dZ_N(t) &= b(Z_N(t)) + \frac{1}{\sqrt{N}}\sigma(Z_N(t))\, dB(t), \\ Z_N(0) &= x_0, \end{cases} \tag{4}$$

where $(B(t))_{t\geq 0}$ is a d- dimensional Brownian motion and σ a $d \times d$ matrix such that

$$\sigma(x)\sigma^t(x) = \Sigma(x). \tag{5}$$

For stochastic differential equations with small noise, an approximation of $Z_N(t)$ is obtained, using (2)-(5), based on the theory of perturbations of dynamical systems (see e.g. Azencott (1982), Freidlin and Wentzell (1978)):

$$\begin{cases} Z_N(t) &= x(t) + \frac{1}{\sqrt{N}}g(t) + \frac{1}{\sqrt{N}}R_N(t), \\ dg(t) &= \nabla_x b(x(t))\; g(t)\, dt + \sigma(x(t))\, dB(t) \quad ; \quad g(0) = 0, \\ \text{with} & \sup_t \|R_N(t)\| \to 0 \;\text{ in probability as } N \to +\infty, \end{cases} \tag{6}$$

where $\nabla_x b(x)$ denotes the matrix $(\frac{\partial b_i}{\partial x_j}(x))_{1\leq i,j\leq d}$. The stochastic differential equation for $g(\cdot)$ defined in (6) can be solved explicitly (see e.g Guy et al. (2014) for details) and its solution is the time inhomogeneous Gaussian process

$$g(t) = \int_0^t \Phi(t, s)\sigma(x(s))\, dB(s), \tag{7}$$

where $\Phi(t, s)$ satisfies $\dfrac{\partial \Phi}{\partial t}(t, s) = \nabla_x b(x(t))\Phi(t, s), \Phi(s, s) = I_d$. Hence, $\Phi(t, s)$ is the $d \times d$ matrix

$$\Phi(t,s) = \exp\left(\int_s^t \nabla_x b(x(u))\, du\right). \tag{8}$$

Using (3) and (7), let us define the Gaussian process $G_N(t)$,

$$G_N(t) = x(t) + \frac{1}{\sqrt{N}} g(t). \tag{9}$$

Consider now the Wasserstein-1 distance on the interval $[0,T]$ between $\mathbb{R}^d$ valued processes U_t, V_t on $[0,T]$. $W_{1,T}(U,V) = \inf \mathbb{E}(\|U-V\|_T)$ where, if $x : [0,T] \to \mathbb{R}^d$, $\|x\|_T = \sup_{0\le t\le T} \|x(t)\|$, and the above infimum is over all the couplings of two processes. According to Britton and Pardoux (2020), Part I, Theorem 2.4.1 page 41, the following holds.

Proposition 1. *For all $T > 0$, the Wasserstein-1 distances on $[0,T]$ between the three processes $(\mathcal{Z}_N(\cdot))$, $(Z_N(\cdot))$ and $(G_N(\cdot))$ defined in* (1),(4),(9) *satisfy, as $N \to \infty$*

$$\sqrt{N} W_{1,T}(\mathcal{Z}_N, Z_N) \to 0, \quad \sqrt{N} W_{1,T}(\mathcal{Z}_N, G_N) \to 0 \quad \textit{and} \ \sqrt{N} W_{1,T}(Z_N, G_N) \to 0.$$

From a statistical point of view, this proposition has important consequences: using that these distances are $o(N^{-1/2})$, we can base our inference method from the observations of the original normalized jump process $(\mathcal{Z}_N)$ as if they followed the diffusion process (Z_N) or the Gaussian process (G_N). Substituting an exact likelihood of a process by the likelihood of another process approximating the first one is often used to derive approximate likelihood or contrasts for stochastic processes. For instance, for discretely observed diffusion processes, the parametric inference is often based on the likelihood of the Euler scheme of the diffusion. We have proved in Guy et al. (2014) that the parametric inference based on (G_N) leads to efficient estimators for the parameters ruling the jump process.
From now on, we will use the approximation of $(\mathcal{Z}_N)$ by the Gaussian process (G_N). Let us consider now a parametric model for the epidemic dynamics. This yields a parametric continuous-time approximate model for the epidemic dynamics, with parameter η containing the parameters entering in the transition rates of the jump process, and therefore in the functions $\beta_\ell(x)$ defined in **H1**, and the initial point x_0 of the ordinary differential equation (ODE) defined in Lemma 1. As mentioned in Section 2.1, the process is however observed at discrete times t_k, where (t_k) is an increasing sequence on $[0,T]$ with $t_0 = 0 < t_1, \cdots, < t_n = T$. We therefore deduce from above a discrete-time representation of the epidemic evolution.
Denote by $\mathcal{F}_t = \sigma(B(s), s \le t)$. Then the following holds.

Proposition 2. *There exists a sequence of independent Gaussian random variables (U_k) such that*

(i) For all k , U_k is $\mathcal{F}_{t_k}$- measurable and U_k independent of $\mathcal{F}_{t_{k-1}}$.

(ii) The process G_N defined in (9) *is an AR(1) process and satisfies, using* (3), (8), *$G_N(0) = x_0$ and for $k \ge 1$,*

$$G_N(t_k) = F_k(\eta) + A_{k-1}(\eta)\, G_N(t_{k-1}) + U_k,$$

where

$$A_{k-1}(\eta) = A(\eta, t_{k-1}) = \Phi(\eta, t_k, t_{k-1}),$$
$$F_k(\eta) = F(\eta, t_k) = x(\eta, t_k) - \Phi(\eta, t_k, t_{k-1})x(\eta, t_{k-1}),$$

and (U_k) *are independent random variables such that*

$$U_k \sim \mathcal{N}_d(0, T_k(\eta)),$$

with

$$T_k(\eta) = \frac{1}{N}\int_{t_{k-1}}^{t_k} \Phi(\eta, t_k, s)\Sigma(\eta, x(\eta, s))\ \Phi^t(\eta, t_k, s)ds.$$

The proof of Proposition 2 is relegated in Appendix. Using now that $\sup_t \|\mathcal{Z}_N(t) - G_N(t)\| = \frac{1}{\sqrt{N}}o_P(1)$, Proposition 2 becomes, setting $X_k := X(t_k) = \mathcal{Z}_N(t_k)$, $X_0 = x_0$ and for $k \geq 1$,

$$X_k = F_k(\eta) + A_{k-1}(\eta)X_{k-1} + U_k. \tag{10}$$

2.3. Approximation of the observation model

Assume now that there are noisy observations $O(t_k)$ of the original jump process $\mathcal{Z}(t)$ (with state space $E = \{0, \ldots, N\}^d$ at discrete times t_k). As mentioned in Section 2.1, it often occurs in practice that not all epidemiological health states are observed. We account for this by introducing a projection operator $B : \mathbb{R}^d \to \mathbb{R}^q$ with $q \leq d$, where $BX(\cdot)$ contains only the coordinates that can be observed. Therefore B is a $d \times q$ matrix whose elements are 0 and 1. For $k = 0, \ldots, n$, define

$$C(t_k) = (C_1(t_k), \ldots, C_q(t_k))^t = B\mathcal{Z}(t_k) \in \{0, \ldots, N\}^q.$$

In a first approach, assume that each component of C is observed with independent reporting rate p_i and measurement errors. Thus, we propose a rather general model for the observations conditionally on $\mathcal{Z}(t)$, for $1 \leq i \leq q$:

$$O_i(t_k) = O_{i,1}(t_k) + O_{i,2}(t_k), \text{ with } O_{i,1}(t_k) \sim \mathcal{B}(C_i(t_k), p_i), \quad O_{i,2}(t_k) \sim \mathcal{N}\left(0, \tau_i^2 C_i(t_k)\right), \tag{11}$$

where, conditionally on $\sigma(\mathcal{Z}(s), 0 \leq s \leq t_k)$, the two variables $O_{i,1}(t_k)$ and $O_{i,2}(t_k)$ are independent. This yields a new parameter of larger dimension, including parameters of both epidemic and observation processes:

$$\theta = (\eta, (p_1, \ldots, p_q), (\tau_1^2, \ldots, \tau_q^2)).$$

Consider now the normalized process $\mathcal{Z}_N(t)$. Then we can define $C_N(t) = B\mathcal{Z}_N(t)$ and associated normalized observations $O_N(t_k) = \frac{1}{N}O(t_k)$. A Gaussian approximation of the observation process has its first and second moments that satisfy

$$\begin{aligned} E(O_{N,i}(t_k)|\mathcal{Z}(t_k)) &= p_i C_{N,i}(t_k), \\ Var(O_{N,i}(t_k)|\mathcal{Z}(t_k)) &= \frac{1}{N}(p_i(1-p_i) + \tau_i^2)C_{N,i}(t_k). \end{aligned}$$

Using now (6) and Proposition 1, we get that

$$C_N(t) = B\mathcal{Z}_N(t) = Bx(\eta, t) + \frac{1}{\sqrt{N}} Bg(\eta, t) + \frac{1}{\sqrt{N}} o_P(1).$$

The Gaussian process $g(\eta, t)$ is uniformly bounded in probability on $[0, T]$ so that

$$Var(O_{N,i}(t_k)|\mathcal{Z}(t_k)) = \frac{1}{N}(p_i(1 - p_i) + \tau_i^2)(Bx(\eta, t_k))_i + O_P(N^{-3/2}).$$

Let us define the two q- dimensional matrices

$$P(\theta) = \mathrm{diag}(p_i)_{1 \le i \le q}, \quad Q_k(\theta) = \frac{1}{N} \mathrm{diag}\left((p_i(1 - p_i) + \tau_i^2)(Bx(\eta, t_k))_i\right), \tag{12}$$

and the $q \times d$ matrix

$$B(\theta) = P(\theta)B.$$

The Gaussian approximations (Y_k) of the observations $O_N(t_k)$ satisfy that, conditionally on $\mathcal{Z}(t_k)$,

$$Y_k = B(\theta)X_k + V_k \text{ with } V_k \sim \mathcal{N}_q(0, Q_k(\theta)), \tag{13}$$

where (V_k) are independent random variables such that, for all k, V_k are independent of $\mathcal{Z}_N(t_k)$.

2.4. Application to the SIR epidemic model

Let us now illustrate the model approximations derived in Sections 2.2 and 2.3 on the simple SIR model introduced in Section 2.1.
The Markov jump process $\mathcal{Z}(t) = (S(t), I(t)))$, $t \ge 0$, is defined in Section 2.2. The parameters ruling the dynamics of the system are

$$\eta = (\lambda, \gamma, x_0) = (\lambda, \gamma, s_0, i_0),$$

which include the starting points $s_0 = S(0)/N$ and $i_0 = I(0)/N$.

Dynamical state model. Let us define the key quantities to derive the appropriate Gaussian process $(G_N(t))$ as defined in (9), including the dependence on η:

$$G_N(t) = x(\eta, t) + \frac{1}{\sqrt{N}} g(\eta, t).$$

The first important element is $x(\eta, t) = (s(\eta, t), i(\eta, t))^t$, solution of the following ordinary differential equation

$$\begin{cases} \frac{ds}{dt}(\eta, t) &= -\lambda s(\eta, t) i(\eta, t), \\ \frac{di}{dt}(\eta, t) &= \lambda s(\eta, t) i(\eta, t) - \gamma i(\eta, t), \\ x_0 &= (s_0, i_0). \end{cases}$$

When there is no ambiguity, we denote by s and i respectively $s(\eta, t)$ and $i(\eta, t)$. Then, to get

$g(\eta, \cdot)$, we need to derive the functions $b(\eta, \cdot)$ and $\Sigma(\eta, \cdot)$ from (2) (see Section 2.2)

$$b(\eta, s, i) = \begin{pmatrix} -\lambda si \\ \lambda si - \gamma i \end{pmatrix}; \quad \Sigma(\eta, s, i) = \begin{pmatrix} \lambda si & -\lambda si \\ -\lambda si & \lambda si + \gamma i \end{pmatrix}, \tag{14}$$

and the Cholesky decomposition of $\Sigma(\eta, \cdot)$:

$$\sigma(\eta, s, i) = \begin{pmatrix} \sqrt{\lambda si} & 0 \\ -\sqrt{\lambda si} & \sqrt{\gamma i} \end{pmatrix}.$$

From (14), we deduce the gradient of b

$$\nabla_x b(\eta, s, i) = \begin{pmatrix} -\lambda i & -\lambda s \\ \lambda i & \lambda s - \gamma \end{pmatrix},$$

and the resolvent matrix (8)

$$\Phi(\eta, t, s) = \exp\left(\int_s^t \nabla_x b(\eta, x(\eta, u)) du \right).$$

Finally, we get

$$g(\eta, t) = \int_0^t \Phi(\eta, t, u) \sigma(\eta, x(\eta, u)) dB(u),$$

where $(B(u))_{u \geq 0}$ is a bi-dimensionnal Brownian motion.

Discrete-time system. For simplicity, we assume a regular sampling: $t_k = k\Delta$, $k = 0, \ldots, n$, $T = n\Delta$. The dependence with respect to Δ is explicitly shown in the equations. The approximate autoregressive model writes, setting $X_k = \mathcal{Z}_N(t_k) = (S_N(k\Delta), I_N(k\Delta))^t$:

$$\begin{cases} X_k & = F_k(\eta, \Delta) + A_{k-1}(\eta, \Delta) X_{k-1} + U_k, \quad \text{where} \\ F_k(\eta, \Delta) & = x(\eta, t_k) - \Phi(\eta, t_k, t_{k-1}) x(\eta, t_{k-1}), \quad A_{k-1}(\eta, \Delta) = \Phi(\eta, t_k, t_{k-1}), \\ U_k \sim \mathcal{N}_2(0, T_k(\eta, \Delta)) & \quad \text{with } T_k(\eta, \Delta) = \frac{1}{N} \int_{t_{k-1}}^{t_k} \Phi(\eta, t_k, s) \Sigma(\eta, x(\eta, s)) \Phi^t(\eta, t_k, s) ds. \end{cases} \tag{15}$$

Observation model. Consider for example that only the infected individuals are observed with reporting and measurement errors. This corresponds to consider in (11):

$$O_1(t_k) = \mathcal{B}(I(t_k), p), \quad O_2(t_k) = \mathcal{N}(0, \tau^2 I(t_k)), \tag{16}$$

hence, the full parameter vector is $\theta = (\lambda, \gamma, s_0, i_0, p, \tau^2)$. To derive (13) from this example, we define the operator $B(\theta) = pB$, where $B : (x_1, x_2)^t \to x_2$ is the projection operator on the infected compartment, and $Q_k(\theta) = \frac{1}{N}(p(1-p) + \tau^2) i(\eta, t_k)$ with Q_k is defined in (12).
By joining (15) with the Gaussian approximate observation model defined above, we get the following discrete-time state-space model

$$\begin{cases} X_k = & F_k(\eta, \Delta) + A_{k-1}(\eta, \Delta) X_{k-1} + U_k, \quad \text{with } U_k \sim \mathcal{N}_2\left(0, T_k(\eta, \Delta)\right), \\ Y_k = & p \begin{pmatrix} 0 & 1 \end{pmatrix} X_k + V_k, \quad \text{with } V_k \sim \mathcal{N}\left(0, \frac{1}{N}(p(1-p) + \tau^2) i(\eta, t_k)\right). \end{cases}$$

3. Parameter estimation by Kalman filtering techniques

3.1. Approximate likelihood inference

The parameters of interest in the generic case are denoted by $\theta = (\eta, (p_1, \ldots, p_q), (\tau_1^2, \ldots, \tau_q^2))$, where η contains the parameters ruling the dynamics and x_0, whereas $(p_1, \ldots, p_q), (\tau_1^2, \ldots, \tau_q^2)$ derive from the reporting and measurements errors in the observations. Our aim is to estimate the unknown parameters θ from observations $y_{n:0} = (y_0, \ldots, y_n)$ obtained at some discrete time points $t_0 < t_1 < \ldots < t_n$. Joining (10) and (13), we get the following discrete-time Gaussian state-space setting that is more convenient for inference:

$$\begin{cases} X_k &= F_k(\eta) + A_{k-1}(\eta)X_{k-1} + U_k, \\ Y_k &= B(\theta)X_k + V_k, \end{cases} \tag{17}$$

where all quantities are explicitly defined in Sections 2.2 and 2.3. Using (17), we propose to estimate θ by maximizing the associated likelihood $L(\cdot; Y_0, \ldots, Y_n)$:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}}\ L(\theta; Y_0, \ldots, Y_n). \tag{18}$$

The log-likelihood of the observations $y_0, \ldots, y_n$ writes as

$$\mathcal{L}(\theta; y_0, \ldots, y_n) = \log f(\theta, y_0) + \sum_{k=1}^{n} \log f_k(\theta; y_k | y_{k-1:0}). \tag{19}$$

Computing $\mathcal{L}(\theta; y_0, \ldots, y_n)$ requires the computation of the two first moments of the Gaussian conditional distributions corresponding to each term $\log f_k(\theta; \ldots)$. It relies on the computation of the predictive distributions $\nu_{k|k-1:0}(\theta; dx) = \mathcal{L}(X_k | y_{k-1:0})$, $k \geq 1$, from which we derive the conditional densities

$$f_k(\theta; y_k | y_{k-1:0}) = \int f(y_k|x)\nu_{k|k-1:0}(\theta; dx).$$

Usually, these conditional distributions are obtained by means of filtering methods, based on the iterative computations of the conditional distributions:

- the *predictive distribution*: $\mathcal{L}(X_k | y_{k-1}, \cdots, y_0) = \nu_{k|k-1:0}(dx)$, $k \geq 1$, with the convention $\nu_{0,0}(dx) = \mathcal{L}(X_0)$,
- the *updating distribution*: $\mathcal{L}(X_k | y_k, \cdots, y_0) = \nu_{k|k:0}(dx)$, $k \geq 0$,
- the *marginal distribution*: $\mathcal{L}(Y_k | y_{k-1}, \cdots, y_0) = \mu_{k|k-1:0}$, $k \geq 1$, with the convention $\mu_{0|0:0}(dx) = \mathcal{L}(Y_0)$.

In the special case of Gaussian state space model and Gaussian noise, all these distributions are Gaussian and therefore characterized by their mean and covariance matrix. Using notations specific to Kalman filtering, let us set

$$\begin{aligned} \mathcal{L}(X_k | y_{k-1}, \cdots, y_0) &= \nu_{k|k-1:0}(dx) = \mathcal{N}_d(\hat{X}_k, \hat{\Xi}_k) \quad (\textit{predictive distribution}). \\ \mathcal{L}(X_k | y_k, \cdots, y_0) &= \nu_{k|k:0}(dx) = \mathcal{N}_d(\bar{X}_k, \bar{T}_k) \quad (\textit{updating distribution}). \\ \mathcal{L}(Y_k | y_{k-1}, \cdots, y_0) &= \mu_{k|k-1:0} = \mathcal{N}_q(\hat{M}_k, \hat{\Omega}_k) \quad (\textit{marginal distribution}). \end{aligned}$$

The Gaussian approximations defined in (10), (13) and (17) enable relying on some specific properties of Gaussian distributions that we recall below.

3.1.1. Preliminary results in the general framework of Kalman filtering

Let $(X_i, i \geq 0)$ be a non centered d-dimensional Gaussian $AR(1)$ process and assume that only q coordinates of (X_i) are observed with Gaussian noise. The computations of the conditional distributions rely on a Kalman filter approach, which derives from the preliminary lemma.

Lemma 2. *Assume that X is a random variable with distribution $\mathcal{N}_d(\xi, T)$ and that, conditionally on X, Y has distribution $\mathcal{N}_q(BX, Q)$. Then, $\mathcal{L}(X|Y)$ is Gaussian $\mathcal{N}_d(\bar{\xi}(y), \bar{T})$ with*

$$\bar{\xi}(y) = \xi + TB^t(BTB^t + Q)^{-1}(y - B\xi); \quad \bar{T} = T - TB^t(BTB^t + Q)^{-1}BT. \tag{20}$$

Remark 1. *Let us stress that Lemma 2 holds even if Q is singular. In particular, the formula holds true when $Q = 0$ and B is a projection operator, i.e. the observations are $Y_k = BX_k$, provided that T is non singular.*

Let us go back to our general setting (X_k, Y_k) defined in (17).

Proposition 3. *Assume that (X_k, Y_k) are defined in* (17)*. Then, $\nu_{k|k-1:0}(dx)$, $\nu_{k|k:0}(dx)$ and $\mu_{k|k-1:0})(dy)$ satisfy, with the initialization $\hat{X}_0 = \xi_0, \hat{\Xi}_0 = T_0$, for $k \geq 0$,*

(i) Prediction : $\nu_{k|k-1:0}(dx) \sim \mathcal{N}_d(\hat{X}_k, \hat{\Xi}_k)$ with
$\hat{X}_k = F_k + A_{k-1}\bar{X}_{k-1}$, $\hat{\Xi}_k = A_{k-1}\bar{T}_{k-1}A_{k-1}^t + T_k$.

(ii) Updating : $\nu_{k|k:0}(dx) \sim \mathcal{N}_d(\bar{X}_k, \bar{T}_k)$ with
$\bar{X}_k = \hat{X}_k + \hat{\Xi}_k B^t(B\hat{\Xi}_k B^t + Q_k)^{-1}(Y_k - B\hat{X}_k)$, $\bar{T}_k = \hat{\Xi}_k - \hat{\Xi}_k B^t(B\hat{\Xi}_k B^t + Q_k)^{-1}B\hat{\Xi}_k$.

(iii) Marginal distribution : $\mu_{k+1|k:0}(dy) \sim \mathcal{N}_q(\hat{M}_{k+1}, \hat{\Omega}_{k+1})$ with
$\hat{M}_{k+1} = B\hat{X}_{k+1}, \quad \hat{\Omega}_{k+1} = B\hat{\Xi}_{k+1}B^t + Q_{k+1}$.

Using some notations from Kalman filtering, we recover a modified version of the Kalman algorithm. Assume that $X_0 \sim \mathcal{N}_d(\xi_0, T_0)$ and that, for all $k \geq 0$, the matrices Γ_k defined below are non singular. Then, setting $\hat{X}_0 = \xi_0, \hat{\Xi}_0 = T_0$, we have

$$\begin{array}{rcll}
\epsilon_{k-1} & = & Y_{k-1} - B\hat{X}_{k-1}, & \textit{(innovation)} \\
\Gamma_{k-1} & = & B\hat{\Xi}_{k-1}B^t + Q_{k-1}, & \textit{(innovation covariance)} \\
H_{k-1} & = & A_{k-1}\hat{\Xi}_{k-1}B^t\Gamma_{k-1}^{-1}, & \textit{(Kalman Gain)} \\
\hat{X}_k & = & F_k + A_{k-1}\hat{X}_{k-1} + H_{k-1}\epsilon_{k-1}, & \textit{(predicted mean state estimation)} \\
\hat{\Xi}_k & = & (A_{k-1} - H_{k-1}B)\hat{\Xi}_{k-1}A_{k-1}^t + T_k. & \textit{(predicted error covariance)}
\end{array}$$

Therefore, the marginal distributions appearing in the computation of the log-likelihood (19) are $\mu_{k+1|k:0}(dy) \sim \mathcal{N}_q(\hat{M}_{k+1}, \hat{\Omega}_{k+1})$ with

$$\hat{M}_{k+1} = B\hat{X}_{k+1}, \quad \hat{\Omega}_{k+1} = B\hat{\Xi}_{k+1}B^t + Q_{k+1}. \tag{21}$$

3.1.2. Recursive computation of the approximate log-likelihood

An important consequence of the previous section is that we can compute (19) based on the recursive computations of the two first moments of the Gaussian distributions corresponding to

each term of the log-likelihood. By explicitly accounting for the dependence on θ of moments given in (21), we obtain:

$$\mathcal{L}(\theta; y_0, \ldots, y_n) = C + \log f(\theta; y_0) - \frac{1}{2}\sum_{k=1}^{n}\left[\log\left(|\hat{\Omega}_k(\theta)|\right) + (y_i - \hat{M}_k(\theta))^t \left(\hat{\Omega}_k(\theta)\right)^{-1} (y_i - \hat{M}_k(\theta))\right],$$

with C a constant (independent of the parameters) and $|A|$ denoting the determinant of the matrix A.
Note that the sampling interval Δ plays an important role in the various key quantities involved in the Kalman recursions (see Appendix A for details).

3.2. Application to the SIR epidemic model

Let us take again the example of SIR epidemics, when only the infected individuals are observed with reporting and measurement errors, considered in Section 2.4.
By assuming an initial distribution $X_0 \sim \mathcal{N}_2(\xi_0, T_0)$, setting $\hat{X}_0 = \xi_0, \hat{\Xi}_0 = T_0$ and applying the algorithm given in Proposition 3, we have, for $k = 0, \ldots, n-1$:

$$\begin{array}{rcll}
\epsilon_{k-1}(\theta) & = & Y_{k-1} - p\hat{I}_{k-1}(\theta), & \text{(scalar)} \\
\Gamma_{k-1}(\theta) & = & p^2(\hat{\Xi}_{k-1}(\theta))_{22} + \frac{1}{N}(p(1-p) + \tau^2)i(\eta, t_{k-1}), & \text{(scalar)} \\
H_{k-1}(\theta) & = & pA_{k-1}(\eta)\hat{\Xi}_{k-1}(\theta)\begin{pmatrix}0\\1\end{pmatrix}\Gamma_{k-1}^{-1}(\theta), & \text{(vector)} \\
\hat{X}_k(\theta) & = & F_k(\eta) + A_{k-1}(\eta)\hat{X}_{k-1}(\theta) + H_{k-1}(\theta)\epsilon_{k-1}(\theta), & \text{(vector)} \\
\hat{\Xi}_k(\theta) & = & \left(A_{k-1}(\eta) - pH_{k-1}(\theta)\begin{pmatrix}0 & 1\end{pmatrix}\right)\hat{\Xi}_{k-1}(\theta)A_{k-1}(\eta)^t + T_k(\eta). & (2\times 2 \text{ matrix})
\end{array}$$

This yields the marginal distributions:

$$\hat{M}_{k+1}(\theta) = p\hat{I}_{k+1}(\theta), \quad \hat{\Omega}_{k+1}(\theta) = p^2\left(\hat{\Xi}_{k+1}(\theta)\right)_{22} + \frac{1}{N}(p(1-p) + \tau^2)i(\eta, t_{k+1})),$$

used to compute the likelihood

$$\mathcal{L}(\theta, y_1, \ldots, y_n) \simeq -\frac{1}{2}\sum_{k=1}^{n}\log\hat{\Omega}_k(\theta) - \frac{1}{2}\sum_{k=1}^{n}\frac{(y_k - \hat{M}_k(\theta))^2}{\hat{\Omega}_k(\theta)}.$$

4. Simulation study

We assessed the performances of our method on simulated SIR epidemics of which only the infectious compartment is observed at discrete time points (see Section 2.4 where the model is fully described).

4.1. Simulation settings

Data simulation. We first simulated SIR dynamics according to the Markov jump process using the Gillespie algorithm (Gillespie (1977)). Only trajectories that did not exhibit early extinction were considered for inference. The theoretical proportion of these trajectories is given by $1 - (\gamma/\lambda)^{I_0}$ (Andersson and Britton (2000)), where I_0 is the number of infectious individuals at time 0. We simulated two cases. First, for the emergent trajectories, the observations were generated by binomial draws from $I(t)$ at $n+1$ discrete time points $t_0 < t_1 < \ldots < t_n$. In (16), this amounts to

considering $\tau = 0$, simulated observations being finally obtained as $O(t_k) = O_1(t_k) \sim \mathcal{B}(I(t_k), p)$. Second, we considered the more general case where observations are $O(t_k) = O_1(t_k) + O_2(t_k)$, with $O_1(t_k) \sim \mathcal{B}(I(t_k), p)$, $O_2(t_k) \sim \mathcal{N}\left(0, \tau^2 I(t_k)\right)$, where the non-zero measurement error τ is an additional parameter to estimate. Figure 2 represents epidemic trajectories corresponding to the different steps of data simulation. The graph illustrates the variability of the stochastic trajectories compared to the deterministic counterpart of the SIR model and the loss of information from the unobservable real dynamics to the observations available for inference. Moreover, the second source of error driven by the measurement error τ seems to have a slight impact on the global observational noise compared to the reporting error. The evolution of the number of susceptible individuals is not shown in Figure 2. From the point of vue of inference, the S compartment is a latent variable, the observations being only available for the infected state.

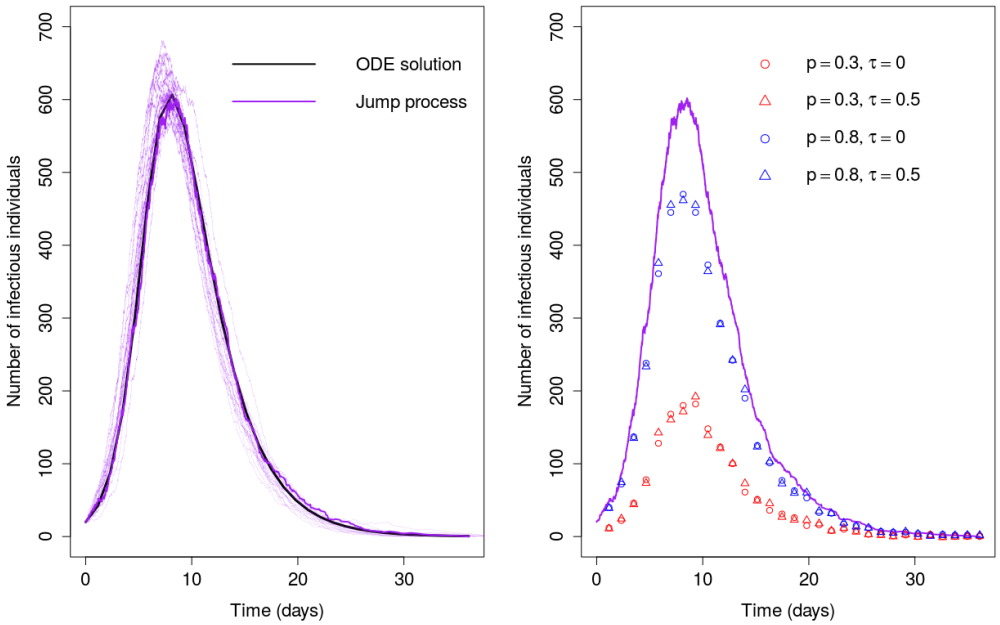

Figure 2: Left panel : ODE solution for the number of infected individuals I (black plain line) and 20 trajectories of the Markov jump process for I (purple lines) when $N = 2000$. Right panel : $n = 30$ observations obtained from a particular trajectory of the jump process (in purple bold in the left panel), as a function of time. The points and triangles stand for observations generated with measurement error terms $\tau = 0$ and $\tau = 0.5$ respectively and the blue and red symbols represent observations generated with $p = 0.8$ and $p = 0.3$ respectively.

Numerical scenarios. We used the following parameter values for the simulation of the epidemics: $\lambda = 1$, $\gamma = 1/3$, and initial starting points $S(0)/N = s_0 = 0.99$, $I(0)/N = i_0 = 0.01$, $R(0) = 0$ (hence with $s_0 + i_0 = 1$). Observations were generated under two scenarios: *i)* high reporting rate $p = 0.8$ and *ii)* low reporting rate $p = 0.3$. Two experiments were considered concerning the measurement error: $\tau = 0$ (experiment one) and $\tau = 0.5$ (experiment two). Scenarios crossing three population sizes ($N \in \{1000, 2000, 10000\}$) and several values for the number of

observations (n) for each epidemic trajectory were also investigated. For each value of N, conditionally on non extinction, 500 SIR epidemic dynamics were simulated. Observations were generated at regularly spaced time points $t_k = k\Delta$ using, for a given scenario, the same value of Δ for each of the 500 epidemics (obtained by dividing the mean epidemic duration over 500 trajectories by a target number of observations n). As the epidemic duration is stochastic, we considered slightly different observation intervals $[0, T]$ for each epidemic and set the value of T as the first time point when the number of infected individuals becomes zero. This generates slightly different numbers of observations per epidemic trajectory.

4.2. Inference: settings, performance comparison and implementation

The unknown parameters to be estimated are either $\theta = (\lambda, \gamma, p, i_0)$ or $\theta = (\lambda, \gamma, p, i_0, \tau)$ according to the experiment. Here, we do not need to estimate s_0 as $s_0 = 1 - i_0$. When $\tau \neq 0$, the observational model used for the two estimation methods was a Gaussian model as the sum of the two sources of noise in data (reporting - Gaussian approximation of a binomial model; measurement - Gaussian model). For each simulated dataset, θ is estimated with our Kalman based estimation method (KM) and with the Maximum Iterated Filtering (MIF) algorithm (Ionides et al. (2006), Ionides et al. (2011), Ionides et al. (2015)), which is widely used in practice for statistical inference of epidemics. The simulation study was performed with the `R` software on a processor Bi-pro Xeon E5-2680, with 2.8 Ghz, 96 Go RAM and 20 cores. MIF estimation is performed with the `mif2` function of the `POMP`-package (King et al. (2017)). We proposed a user-friendly code available on the RunMyCode website (see Appendix F for details).

Let us make some remarks on the algorithms and their practical implementations.
Regardless of the method used for the estimation, the maximisation of the log-likelihood necessitates considering several constraints: (i) strict positivity of λ, γ, i_0, (ii) $s_0 + i_0 = 1$ (or $s_0 + i_0 \leq 1$ in the general case) and (iii) $0 < p \leq 1$. To facilitate the optimisation, a different parameterization was implemented: $\lambda = \exp(\mu_1)$, $\gamma = \exp(\mu_2)$, $p = (1 + \exp(\mu_3))^{-1}$, $i_0 = (1 + \exp(\mu_4))^{-1}$ where $\mu_1, \mu_2, \mu_3, \mu_4 \in \mathbb{R}$. With no constraint on the new set of parameters, the numerical optimization was more stable in practice.
The approximated log-likelihood given by Kalman filtering techniques cannot be maximized explicitly. We then use the Nelder-Mead method implemented in the `optim` function of the `R` software, which requires to provide some initial values for the unknown parameters. According to the amount of information available in the observations, the result of the optimization is more or less sensitive to these initial points. The same problem can occur for the MIF algorithm. The dependence on the initialization can be circumvented by trying different starting values (10 in the present case) and choosing the maximum value for the log-likelihood among them. The starting parameter values for the maximization algorithm were uniformly drawn from a hypercube encompassing the true values.
When the time intervals between observations Δ are large (which often occurs for low values of n), we computed the resolvent matrix following (A.1) in order to obtain the approximated log-likelihood with Kalman filtering techniques.
MIF, based on particle filtering, returns an estimate of the log-likelihood of the observations by using re-sampling techniques. The parameter space is investigated by randomly perturbing the parameters of interest at each iteration, the amplitude of the perturbation decreasing as the iterations progress. The MIF algorithm has a complexity $O(JM)$, where J and M are respectively the number of particles and the number of iterations. Running MIF requires specifying several

tuning parameters. For the present study, the best results were obtained by using $M = 100$ iterations, $J = 500$ particles, standard deviations `rw.sd` equal to 0.2 for the random walk for each parameter and a cooling of the perturbations of `cooling.fraction.50=0.05` in the `POMP`-package (we drew inspiration from Stocks (2017) for this choice of tuning parameters).
Concerning implementation issues, from our experience, the tuning of the MIF algorithm (number of particles, number of iterations, etc.) can greatly affect the quality of the estimates. In particular, it seems that there is an important interplay between the tuning parameters and the initialization values of the model parameter to be inferred. In comparison, our method has only one main practical calibration parameter. In the filtering step, it is necessary to initialize the covariance matrix (*i.e.* T_0 in Section 3.2) of the state variables conditionally to the observations but it seems that this initialization does not have a noticeable influence on the estimates accuracy.

4.3. Point estimates and standard deviations for key model parameters θ

4.3.1. Numerical results for the first experiment ($\tau = 0$)

Three different target values for sample sizes were considered: $n = 10$, $n = 30$ and $n = 100$. Tables 1 and 2 respectively display the results for the high reporting scenario ($p = 0.8$) and the low reporting scenario ($p = 0.3$). Each table compares estimates obtained with KM and MIF. For each parameter and each estimation method, the reported values are the mean of the 500 parameter estimates and their standard deviations in brackets.
The results show that, irrespective to the reporting rate p, when the population size N and the number of observations n per epidemic increase, the bias and the standard error of the estimates obtained decrease whatever the method used for inference. For a given (N, n), the estimation bias is higher when the reporting rate is low ($p^* = 0.3$, where * designates here the true value). This can be partly related to the fact that the information contained in the data deteriorates as p^* decreases. Both methods provide estimates with comparable levels of accuracy.

Table 1: First experiment ($\tau = 0$). Estimation of $\theta = (\lambda, \gamma, p, i_0)$ under the constraint $s_0 + i_0 = 1$ in setting 1 with true parameter values $(\lambda^*, \gamma^*, p^*, i_0^*)$=$(1, 1/3, 0.8, 0.01)$. For each combination of (N, n) and for each model parameter, point estimates and standard deviations are calculated as the mean of the 500 individual estimates and their standard deviations (in brackets) obtained by our Kalman-based method (KM) and Maximum Iterated Filtering (MIF). The reported values for the number of observations n correspond to the average over the 500 trajectories, with the min and the max in brackets.

	$N = 1000$			$N = 2000$			$N = 10000$		
	$n = 11$	$n = 31$	$n = 101$	$n = 11$	$n = 31$	$n = 102$	$n = 10$	$n = 30$	$n = 100$
	$(7, 18)$	$(21, 51)$	$(68, 168)$	$(8, 19)$	$(23, 55)$	$(75, 179)$	$(8, 15)$	$(25, 44)$	$(81, 143)$
$\lambda^* = 1$									
KM	1.01	0.99	0.99	1.02	1.00	1.00	1.02	1.00	1.00
	(0.09)	(0.08)	(0.07)	(0.06)	(0.06)	(0.06)	(0.03)	(0.03)	(0.03)
MIF	1.02	0.99	1.00	1.01	1.00	1.01	1.01	1.00	1.00
	(0.07)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.02)	(0.02)	(0.02)
$\gamma^* = 1/3$									
KM	0.30	0.31	0.33	0.31	0.32	0.33	0.32	0.33	0.34
	(0.03)	(0.04)	(0.03)	(0.03)	(0.04)	(0.03)	(0.02)	(0.02)	(0.02)
MIF	0.32	0.31	0.34	0.32	0.32	0.34	0.33	0.32	0.34
	(0.04)	(0.04)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
$p^* = 0.8$									
KM	0.70	0.73	0.79	0.73	0.75	0.79	0.77	0.78	0.82
	(0.10)	(0.11)	(0.06)	(0.08)	(0.11)	(0.07)	(0.05)	(0.06)	(0.05)
MIF	0.75	0.74	0.80	0.77	0.74	0.80	0.78	0.74	0.81
	(0.11)	(0.09)	(0.04)	(0.09)	(0.08)	(0.05)	(0.06)	(0.04)	(0.04)
$i_0^* = 0.01$									
KM	0.011	0.016	0.012	0.010	0.013	0.011	0.010	0.010	0.010
	(0.005)	(0.008)	(0.006)	(0.003)	(0.006)	(0.005)	(0.001)	(0.002)	(0.003)
MIF	0.011	0.012	0.011	0.011	0.012	0.011	0.010	0.011	0.010
	(0.005)	(0.004)	(0.002)	(0.003)	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)

Table 2: First experiment ($\tau = 0$). Estimation of $\theta = (\lambda, \gamma, p, i_0)$ under the constraint $s_0 + i_0 = 1$ in setting 2 with true parameter values $(\lambda^*, \gamma^*, p^*, i_0^*)$=$(1, 1/3, 0.3, 0.01)$. For each combination of (N, n) and for each model parameter, point estimates and standard deviations are calculated as the mean of the 500 individual estimates and their standard deviations (in brackets) obtained by KM and MIF. The reported values for the number of observations n correspond to the average over the 500 trajectories, with the min and the max in brackets.

	$N = 1000$			$N = 2000$			$N = 10000$		
	$n = 11$	$n = 31$	$n = 101$	$n = 11$	$n = 31$	$n = 102$	$n = 10$	$n = 30$	$n = 100$
	$(7, 18)$	$(21, 51)$	$(68, 168)$	$(8, 19)$	$(23, 55)$	$(75, 179)$	$(8, 15)$	$(25, 44)$	$(81, 143)$
$\lambda^* = 1$									
KM	1.01	1.04	1.00	1.00	1.02	1.01	0.99	1.02	1.00
	(0.10)	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)	(0.03)	(0.03)	(0.03)
MIF	1.02	1.07	1.01	0.99	1.03	1.02	0.98	1.01	1.00
	(0.09)	(0.07)	(0.06)	(0.06)	(0.04)	(0.05)	(0.03)	(0.02)	(0.02)
$\gamma^* = 1/3$									
KM	0.26	0.30	0.32	0.28	0.32	0.32	0.31	0.33	0.34
	(0.03)	(0.05)	(0.05)	(0.03)	(0.05)	(0.05)	(0.02)	(0.02)	(0.03)
MIF	0.27	0.30	0.31	0.28	0.32	0.32	0.31	0.34	0.33
	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
$p^* = 0.3$									
KM	0.21	0.26	0.29	0.23	0.29	0.29	0.27	0.30	0.30
	(0.03)	(0.05)	(0.05)	(0.03)	(0.05)	(0.05)	(0.02)	(0.03)	(0.03)
MIF	0.22	0.26	0.27	0.23	0.28	0.28	0.27	0.30	0.29
	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
$i_0^* = 0.01$									
KM	0.010	0.007	0.010	0.012	0.009	0.011	0.011	0.010	0.011
	(0.006)	(0.004)	(0.006)	(0.004)	(0.004)	(0.004)	(0.002)	(0.002)	(0.002)
MIF	0.012	0.008	0.009	0.013	0.009	0.009	0.012	0.010	0.010
	(0.007)	(0.004)	(0.003)	(0.004)	(0.003)	(0.002)	(0.002)	(0.001)	(0.001)

The estimations are less computationally demanding and they require less algorithmic tuning with the Kalman filtering approach. This numerical study was also performed for a second set of parameter values ($\lambda = 0.6$, $\gamma = 0.4$, $i_0 = 0.01$), under the constraint $s_0 + i_0 = 1$ and for $p = 0.8$ and $p = 0.3$, that naturally leads to greater variability between simulated trajectories. The results are provided in Appendix E for comparative purposes.

4.3.2. Numerical results for the second experiment ($\tau \neq 0$)

Here, we present the estimation results when the simulated observations are obtained with a non-zero measurement error τ to be estimated. As noticed in Stocks et al. (2018), the initial conditions of the system are difficult to estimate, and usually set at plausible values. Consequently, we distinguish two situations where (i) i_0 is unknown and estimated; (ii) i_0 is known and fixed.

Unknown starting point i_0. Five different target values for sample sizes were considered: $n = 10$, $n = 30$, $n = 100$, $n = 500$ and $n = 1000$. The unknown parameters to be estimated are $\theta = (\lambda, \gamma, p, i_0, \tau)$ under the constraint $s_0 + i_0 = 1$. For the sake of clarity, we do not show the results when $N = 2000$ and $p = 0.3$. The results are displayed in Table 3.

Table 3: Second experiment ($\tau \neq 0$). Estimation of $\theta = (\lambda, \gamma, p, i_0, \tau)$ under the constraint $s_0 + i_0 = 1$ in setting 1 with true parameter values $(\lambda^*, \gamma^*, p^*, i_0^*, \tau^*)$=$(1, 1/3, 0.8, 0.01, 0.5)$. For each combination of (N, n) and for each model parameter, point estimates and standard deviations are calculated as the mean of the 500 individual estimates and their standard deviations (in brackets) obtained by our Kalman-based method and the MIF algorithm. The reported values for the number of observations n correspond to the average over the 500 trajectories, with the min and the max in brackets.

	$N = 1000$					$N = 10000$				
	$n = 11$	$n = 31$	$n = 101$	$n = 501$	$n = 1001$	$n = 10$	$n = 30$	$n = 100$	$n = 499$	$n = 998$
	$(7, 18)$	$(21, 51)$	$(68, 168)$	$(338, 833)$	$(676, 1665)$	$(8, 15)$	$(25, 44)$	$(81, 143)$	$(406, 716)$	$(811, 1430)$
$\lambda^* = 1$										
KM	0.99	0.98	0.98	0.97	0.99	1.01	0.99	0.99	0.98	0.99
	(0.10)	(0.08)	(0.07)	(0.08)	(0.08)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)
MIF	1.02	1.00	1.02	1.01	1.00	1.01	1.00	1.01	1.00	1.01
	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
$\gamma^* = 1/3$										
KM	0.29	0.30	0.31	0.32	0.32	0.32	0.32	0.33	0.32	0.33
	(0.03)	(0.05)	(0.05)	(0.06)	(0.07)	(0.02)	(0.02)	(0.02)	(0.03)	(0.04)
MIF	0.30	0.30	0.31	0.32	0.33	0.32	0.31	0.33	0.33	0.34
	(0.03)	(0.04)	(0.04)	(0.03)	(0.04)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
$p^* = 0.8$										
KM	0.67	0.72	0.74	0.76	0.75	0.75	0.76	0.79	0.77	0.80
	(0.09)	(0.15)	(0.12)	(0.13)	(0.15)	(0.05)	(0.07)	(0.07)	(0.07)	(0.11)
MIF	0.70	0.70	0.74	0.75	0.78	0.75	0.72	0.78	0.77	0.82
	(0.09)	(0.09)	(0.10)	(0.08)	(0.11)	(0.05)	(0.04)	(0.05)	(0.05)	(0.06)
$i_0^* = 0.01$										
KM	0.011	0.014	0.016	0.014	0.014	0.010	0.011	0.010	0.011	0.009
	(0.005)	(0.006)	(0.006)	(0.005)	(0.010)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
MIF	0.011	0.012	0.012	0.012	0.011	0.011	0.011	0.011	0.011	0.010
	(0.004)	(0.004)	(0.003)	(0.002)	(0.003)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)
$\tau^* = 0.5$										
KM	0.05	0.48	0.48	0.46	0.44	0.05	0.47	0.42	0.43	0.51
	(0.16)	(0.22)	(0.13)	(0.17)	(0.21)	(0.16)	(0.19)	(0.08)	(0.09)	(0.13)
MIF	0.48	0.52	0.46	0.48	0.49	0.60	0.54	0.39	0.46	0.53
	(0.21)	(0.15)	(0.12)	(0.09)	(0.13)	(0.21)	(0.14)	(0.09)	(0.06)	(0.07)

As in the first experiment where $\tau = 0$, the results show that the estimations provided by KM and MIF are of the same order of accuracy. The pattern concerning the bias and the standard error observed in the case $\tau = 0$ also occurs when $\tau = 0.5$ is estimated, *i.e.* bias decreasing and accuracy increasing when N and n increase. We remark that the estimation is more difficult, inducing larger bias, when the measurement error τ is unknown, even for a quite large number of observations $n \approx 100$. Consider for example $N = 1000$, $n = 101$ and $p^* = 0.8$. The point estimate value of p obtained by KM with $\tau = 0$ (cf. Table 1) and $\tau \neq 0$ (cf. Table 3) is respectively 0.79 and 0.74. This is more marked for the second set of parameters values ($\lambda = 0.6$ and $\gamma = 0.4$) which is set out in Appendix E and which induces more variability between epidemics. For $N = 1000$, $n = 99$ and $p^* = 0.8$, comparing the results in Tables E.6 and E.8 shows that $\hat{p}$ passes from 0.75 to 0.66 when $\tau = 0.5$ unknown. Higher frequency observations of the epidemics lead to more satisfactory estimations: considering $n = 998$ when $\tau = 0.5$ unknown leading to $\hat{p} = 0.78$. The estimates obtained with MIF have the same behaviour. In summary, when the measurement error τ is non-zero and estimated, a higher value of the number of observations is needed in order to obtain estimations without bias for both Kalman-based and MIF methods.

Known starting point i_0. The unknown parameters to be estimated are $\theta = (\lambda, \gamma, p, \tau)$. Tables 4 and 5 respectively display the results for the high reporting scenario ($p = 0.8$) and the low reporting scenario ($p = 0.3$).

Table 4: Second experiment ($\tau \neq 0$). Estimation of $\theta = (\lambda, \gamma, p, \tau)$ with $s_0 = 0.99$ and $i_0 = 0.01$ known in setting 1 with true parameter values $(\lambda^*, \gamma^*, p^*, \tau^*)$=$(1, 1/3, 0.8, 0.5)$. For each combination of (N, n) and for each model parameter, point estimates and standard deviations are calculated as the mean of the 500 individual estimates and their standard deviations (in brackets) obtained by KM and MIF. The reported values for the number of observations n correspond to the average over the 500 trajectories, with the min and the max in brackets.

	$N = 1000$			$N = 2000$			$N = 10000$		
	$n = 11$	$n = 31$	$n = 101$	$n = 11$	$n = 31$	$n = 102$	$n = 10$	$n = 30$	$n = 100$
	$(7, 18)$	$(21, 51)$	$(68, 168)$	$(8, 19)$	$(23, 55)$	$(75, 179)$	$(8, 15)$	$(25, 44)$	$(81, 143)$
$\lambda^* = 1$									
KM	1.04	1.01	1.01	1.03	0.98	1.01	1.01	0.99	1.00
	(0.12)	(0.08)	(0.08)	(0.08)	(0.07)	(0.07)	(0.04)	(0.03)	(0.03)
MIF	1.03	1.01	1.02	1.02	1.01	1.02	1.02	1.00	1.01
	(0.08)	(0.07)	(0.07)	(0.05)	(0.05)	(0.05)	(0.02)	(0.02)	(0.02)
$\gamma^* = 1/3$									
KM	0.29	0.31	0.32	0.29	0.30	0.31	0.31	0.32	0.33
	(0.04)	(0.06)	(0.05)	(0.03)	(0.04)	(0.04)	(0.02)	(0.02)	(0.02)
MIF	0.30	0.31	0.33	0.31	0.31	0.32	0.32	0.32	0.33
	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
$p^* = 0.8$									
KM	0.69	0.76	0.76	0.71	0.71	0.75	0.74	0.76	0.79
	(0.11)	(0.16)	(0.13)	(0.09)	(0.12)	(0.10)	(0.05)	(0.06)	(0.06)
MIF	0.70	0.72	0.78	0.73	0.72	0.76	0.75	0.73	0.79
	(0.09)	(0.10)	(0.10)	(0.08)	(0.07)	(0.08)	(0.06)	(0.05)	(0.04)
$\tau^* = 0.5$									
KM	0.11	0.54	0.52	0.08	0.34	0.50	0.11	0.50	0.43
	(0.23)	(0.23)	(0.14)	(0.22)	(0.22)	(0.10)	(0.26)	(0.18)	(0.08)
MIF	0.49	0.54	0.50	0.49	0.48	0.48	0.60	0.56	0.42
	(0.22)	(0.15)	(0.11)	(0.22)	(0.14)	(0.09)	(0.23)	(0.13)	(0.07)

Table 5: Second experiment ($\tau \neq 0$). Estimation of $\theta = (\lambda, \gamma, p, \tau)$ with $s_0 = 0.99$ and $i_0 = 0.01$ known in setting 2 with true parameter values $(\lambda^*, \gamma^*, p^*, \tau^*)$=$(1, 1/3, 0.3, 0.5)$. For each combination of (N, n) and for each model parameter, point estimates and standard deviations are calculated as the mean of the 500 individual estimates and their standard deviations (in brackets) obtained by KM and MIF. The reported values for the number of observations n correspond to the average over the 500 trajectories, with the min and the max in brackets.

	$N = 1000$			$N = 2000$			$N = 10000$		
	$n = 11$	$n = 31$	$n = 101$	$n = 11$	$n = 31$	$n = 102$	$n = 10$	$n = 30$	$n = 100$
	$(7, 18)$	$(21, 51)$	$(68, 168)$	$(8, 19)$	$(23, 55)$	$(75, 179)$	$(8, 15)$	$(25, 44)$	$(81, 143)$
$\lambda^* = 1$									
KM	0.99	1.01	1.05	1.03	0.99	1.01	1.01	0.99	1.01
	(0.14)	(0.09)	(0.08)	(0.11)	(0.07)	(0.07)	(0.04)	(0.03)	(0.03)
MIF	1.05	1.06	1.05	1.06	1.02	1.03	1.02	1.01	1.01
	(0.14)	(0.08)	(0.07)	(0.10)	(0.05)	(0.05)	(0.04)	(0.02)	(0.02)
$\gamma^* = 1/3$									
KM	0.24	0.28	0.29	0.26	0.30	0.29	0.28	0.32	0.34
	(0.06)	(0.05)	(0.05)	(0.07)	(0.04)	(0.04)	(0.03)	(0.02)	(0.02)
MIF	0.23	0.29	0.29	0.24	0.30	0.30	0.28	0.32	0.34
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
$p^* = 0.3$									
KM	0.20	0.25	0.26	0.22	0.27	0.26	0.24	0.29	0.31
	(0.07)	(0.05)	(0.05)	(0.07)	(0.05)	(0.04)	(0.03)	(0.03)	(0.03)
MIF	0.19	0.25	0.25	0.20	0.27	0.27	0.24	0.29	0.30
	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
$\tau^* = 0.5$									
KM	0.15	0.16	0.44	0.17	0.12	0.32	0.08	0.20	0.52
	(0.15)	(0.12)	(0.06)	(0.18)	(0.12)	(0.06)	(0.16)	(0.15)	(0.05)
MIF	0.41	0.26	0.44	0.45	0.24	0.36	0.50	0.30	0.50
	(0.12)	(0.10)	(0.04)	(0.12)	(0.10)	(0.06)	(0.13)	(0.09)	(0.04)

It seems that the influence of knowing or not knowing the initial condition i_0 is different according to the values of the parameters used to simulate the data. For the setting where $\lambda = 1$ and $\gamma = 1/3$, comparing Tables 3 and 4 does not exhibit major differences between estimates. On the contrary, the impact of knowing or not knowing the initial condition i_0 is more visible when considering $\lambda = 0.6$ and $\gamma = 0.4$ (see Appendix E). Tables E.8 and E.9 show that the quality of the estimates is deteriorating when i_0 is unknown leading in particular to more significant biases. For $N = 10000$, $n = 101$ and $p^* = 0.8$, $\hat{p}$ passes from 0.77 when i_0 is known to 0.65 when i_0 is unknown. Once again, higher frequency observations of the epidemics lead to more satisfactory estimations (see Table E.8). Tables 4 and 5 suggest that the estimation bias obtained for the measurement error τ increases when p^* decreases.

4.3.3. Additional comments

In the simulation study, we also considered cases where only the susceptible are observed (not shown here). We noticed that the estimations provided by our Kalman-based method and the MIF algorithm are more accurate by considering the S values rather than the I values. As the S values are several orders of magnitude higher than the I values, a plausible explanation is that the observation noise (due to imperfect reporting and measurement errors) has a lower impact on the S values.

As regards the computation times of both methods, they are sensitive to the number of observations n per epidemics: the computation time increases linearly with n. Concerning the population size N, only the computation time for MIF-based inference is increased when N increases, our method being insensitive to it. As an example, for the scenario with $N = 10000$, $n = 30$ and

$p = 0.8$ (which corresponds to Table 1), the average computation time for a single estimation (*i.e.* a single trajectory) is 31 seconds with KM and 97 seconds with the MIF algorithm. For $n = 100$, the average computation times are 81 and 147 seconds for the KM and the MIF algorithm respectively.

4.4. Confidence interval estimates based on profile likelihood

Following other authors (see Ionides et al. (2017) for instance), we provided profile-likelihood confidence intervals of estimated parameters, for which we briefly recall the principle. Let us denote a general parameters vector $\psi = (\psi_1, \psi_2)$ where $\psi_1 \in \mathbb{R}$ is the parameter of interest and ψ_2 contains the remaining parameters. The profile log-likelihood of ψ_1 is built by maximizing the approximate log-likelihood function (proposed in Section 3) over ψ_2, for fixed values of ψ_1: $\mathcal{L}_{profile}(\psi_1) = \max_{\psi_2} \mathcal{L}(\psi_1, \psi_2)$. A 95% confidence interval for ψ_1 is given by:

$$\{\psi_1 : \mathcal{L}_{profile}(\hat{\psi}) - \mathcal{L}_{profile}(\psi_1)\} < 1.92, \tag{22}$$

where $\hat{\psi}$ is the maximum approximate likelihood estimator (see (18)). The threshold 1.92 comes from Wilks' theorem and corresponds to the quantile of order 0.95 of the χ^2 distribution with 1 degree of freedom.

As an illustrative example, 95% profile likelihood confidence intervals were constructed for the key epidemic parameters λ and γ on two particular trajectories of SIR simulated dynamics in the first experiment ($\tau = 0$). A graphical representation is provided in Figure 3 for parameter λ and in Figure 4 for parameter γ. The first confidence interval (left panel of both Figures) is obtained with a sample of $n = 30$ observation of an SIR epidemic for a population of size $N = 2000$ with reporting rate $p = 0.3$. The second confidence interval (right panel of each Figures) is obtained with a sample of $n = 100$ observation of an SIR epidemic for a population of size $N = 10000$ with reporting rate $p = 0.8$. For each of the two parameters (playing the role of ψ_1 in (22)), 20 values were considered in a relevant interval containing the point estimate. For each of the 20 values of the parameter of interest, the remaining parameters (playing the role of ψ_2 in (22)), on which the likelihood is optimized (corresponding to $\mathcal{L}_{profile}(\psi_1)$ in (22))), were randomly initialized, with 10 different initialization values, the best one being stored. The 20 values of maximum log-likelihood were reported on a graph, relied by a smoothing curve. The two vertical lines, going through the intersection of this curve with the horizontal line at the ordinate equal to the maximum log-likelihood for all parameters minus 1.92 (cf. equation (22)), determine the abscissa for the CI95%. Based on Figures 3 and 4, we see that the width of the confidence intervals, $\text{CI95\%}(\lambda) = [0.96, 1.10]$ and $\text{CI95\%}(\gamma) = [0.31, 0.48]$, is naturally greater in the case where $N = 2000$, $n = 30$ and $p = 0.3$ (which is a more difficult case for performing estimations, due to an increased stochasticity of epidemic trajectories and a significant noise in observations) than for $N = 10000$, $n = 100$ and $p = 0.8$ (a much more tractable case with small variability among trajectories and a low amount of noise in observations), $\text{CI95\%}(\lambda) = [0.95, 1.00]$ and $\text{CI95\%}(\gamma) = [0.33, 0.36]$.

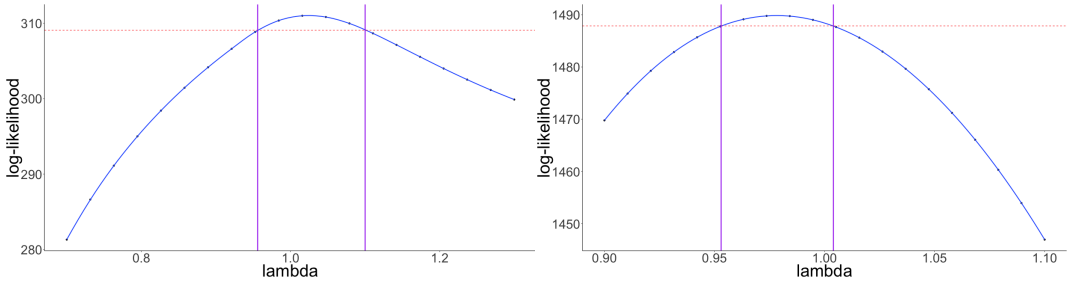

Figure 3: Profile likelihood and confidence intervals (CI95%) for λ. Left panel: data simulated with $N = 2000$, $n = 30$ and $p = 0.3$; the true value $\lambda^* = 1$, the point estimate $\hat{\lambda} = 1.02$ and CI95% $= [0.96, 1.10]$. Right panel : data simulated with $N = 10000$, $n = 100$ and $p = 0.8$; the true value $\lambda^* = 1$, the point estimate $\hat{\lambda} = 1.00$ and CI95% $= [0.95, 1.00]$.

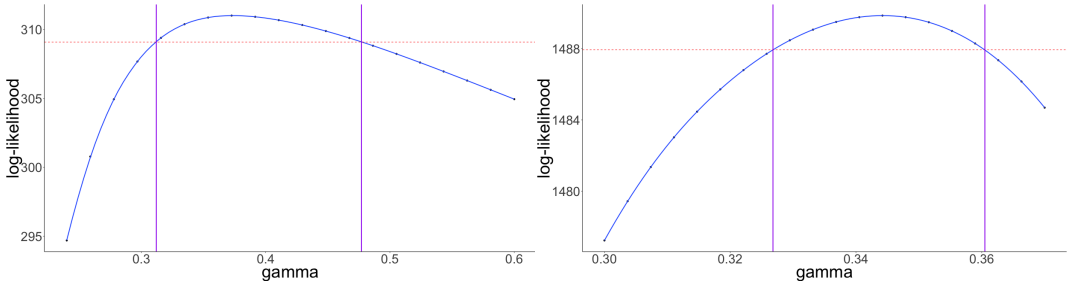

Figure 4: Profile likelihood and confidence intervals (CI95%) for γ. Left panel: data simulated with $N = 2000$, $n = 30$ and $p = 0.3$; the true value $\gamma^* = 1/3$, the point estimate $\hat{\gamma} = 0.32$ and CI95% $= [0.31, 0.48]$. Right panel: data simulated with $N = 10000$, $n = 100$ and $p = 0.8$; the true value $\gamma^* = 1/3$, the point estimate $\hat{\gamma} = 0.34$ and CI95% $= [0.33, 0.36]$.

5. Application to real data

We applied our inference method on the data from an influenza outbreak that occurred in January 1978 in a boarding school in the North of England (Anonymous (1978)), with $N = 763$. The observations correspond to the daily number of infectious boys over 14 days ($n = 14$). It is known that the epidemics started due to a single infectious student. Here also, we assumed that the epidemic dynamics followed an SIR model. Hence, $S(0) = 762$ and $I(0) = 1$ and the parameters to be estimated are the epidemic parameters (λ, γ), the reporting rate p and the parameter τ related to observational noise.

Estimations were performed with both KM and MIF. For the MIF method, we used the same tuning parameters values than those chosen in the simulation study. Both series of results were graphically assessed by post-predictive checks. For that purpose, Markov jump processes of the SIR model were simulated using each set of parameter estimates. We kept 1000 trajectories that did not exhibit early extinction according to the theoretical criterion used in section 4.1. From these 1000 trajectories, we then generated equally spaced observations with $n = 14$. Empirical mean, 5th, 50th and 95th percentiles were extracted time by time and superimposed to the real data (Figure 5).

The following estimations were obtained, with the profile-likelihood-based confidence intervals (CI95%) provided in brackets:

- $\hat{\lambda}_{\mathrm{KM}} = 1.72\ [1.61, 1.83]$; $\hat{\gamma}_{\mathrm{KM}} = 0.48\ [0.43, 0.52]$; $\hat{p}_{\mathrm{KM}} = 1.00\ [0.92, 1.00]$; $\hat{\tau}_{\mathrm{KM}} = 0.91\ [0.42, 1.62]$ with KM,
- $\hat{\lambda}_{\mathrm{MIF}} = 1.85\ [1.62, 2.15]$; $\hat{\gamma}_{\mathrm{MIF}} = 0.47\ [0.39, 0.54]$; $\hat{p}_{\mathrm{MIF}} = 0.97\ [0.84, 1.00]$; $\hat{\tau}_{\mathrm{MIF}} = 1.58\ [0.80, 2.80]$ with MIF.

The estimated values for λ, γ and p are close for both methods, but the estimation values for τ are rather different. The confidence intervals provided by the MIF method are larger than those obtained by our Kalman-based method, but this could be due to a non optimal tuning in case of MIF. Moreover, we notice that the confidence interval of τ is particularly large for both methods, which is in agreement with the fact that a moderate number of observations is needed in order to properly estimate τ (as showed in simulation analysis). The post-predictive check indicates (Figure 5) that both methods provide estimations and hence predictions that are consistent with the data. Estimation takes 22.7 seconds with our method whereas it takes 46.5 seconds with MIF.

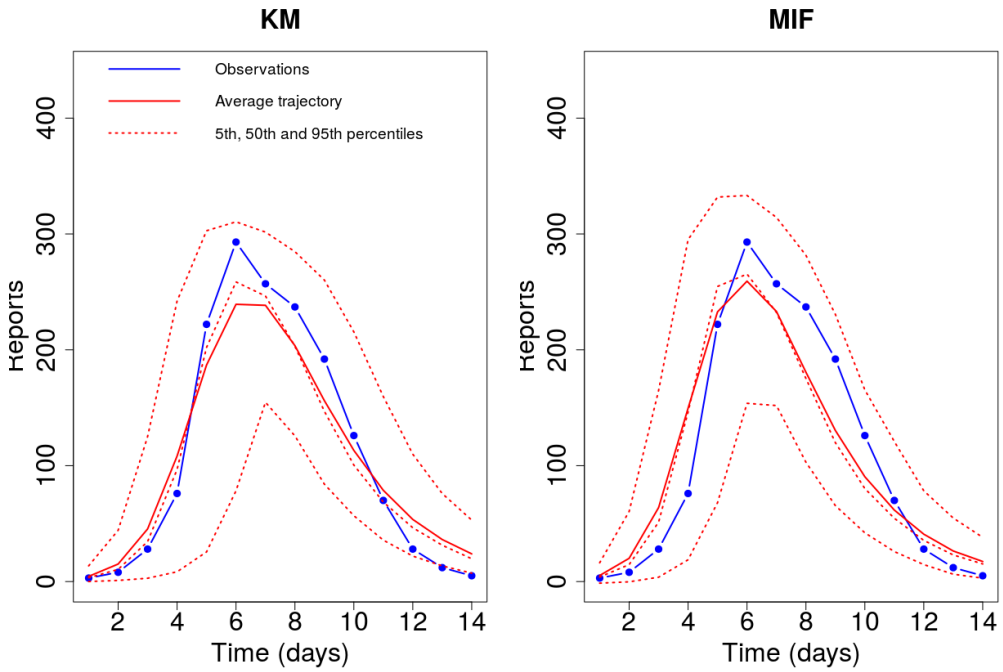

Figure 5: Post-predictive checks for the Kalman-based (KM, left panel) and the Maximum Iterated Filtering (MIF, right panel) estimations. In blue: observations (numbers of infectious boys). Red solid line: average trajectory over 1000 Markov jump processes from the estimated model. Red dotted lines: 5th, 50th and 95th percentiles.

6. Discussion

In this paper we propose a generic and easily practicable inference method for discrete, partially and noisily observed continuous-time epidemic models from time-series data. We derived

a Gaussian approximation of the epidemic density-dependent Markovian jump process underlying epidemic dynamics, through a diffusion based approach, and a Gaussian approximation of the model of observations. This two-level Gaussian approximation allowed us to develop an inference method based on Kalman filtering for the calculation of the likelihood, to estimate key epidemic parameters (such as transmission and recovery rates), initial state of the system (number of susceptible and infectious) and parameters of observation model (such as the reporting rate) from parsimonious and noisy data (proportion of infectious over time).

The performances of the estimators obtained with the Kalman-based method were investigated on simulated data, under various scenarios, with respect to the parameter values of epidemic and observation processes, the population size (N), the number of observations (n) and the nature of data (number of susceptibles S or infectious I over time). The performances, in terms of bias but even more of accuracy, were improved when increasing N and especially n, becoming satisfactory for realistic observation design (e.g. $n = 30$ which corresponds in our case to one observation per day or every two days) and moderate community size ($N = 2000$). The influence of N and n is less pronounced when data are less incomplete (here in the case where p, the proportion of available data, corresponding to the reporting rate, was equal to 0.8 and τ, corresponding to the measurement error, is zero). The estimation is more challenging when the measurement error τ is unknown . In this case, higher frequency observations are needed in order to obtain more accurate estimations. When in addition to a non-zero measurement error, the starting point i_0 is unknown, the quality of the estimates could be deteriorated in some cases. Similar performances were observed irrespective to the data type (when observations are sampled from S instead of I; results not shown). In addition, our method seems to be little impacted by tuning aspects. Indeed, the only obvious tuning parameter, concerning the initialization of the covariance matrix of the state variables conditionally upon the observations, in the filtering step, does not seem to influence the estimation accuracy. Besides simulated data, our method provides very satisfying estimations when applied to real data, from an influenza outbreak in a Britain boarding school in 1978, as revealed by the post-predictive check showing consistency with data. These good performances are all the more noteworthy as the numerical scenario involves certain difficulties (low N and n).

Estimations obtained with the KM were compared to those obtained using MIF (Ionides et al. (2011), King et al. (2017)). The MIF algorithm is efficient in terms of inference performances, but computationally expensive and uses tuning parameters (number of particles, number of iterations, etc.) that are crucial for the successful functioning of the procedure. Importantly, our method does not require such specific computational calibration and its computation time is lower.

As a limitation of our method, we noticed that joint estimation of parameters of epidemic and observation models (λ, γ, p) along with the initial conditions of the underlying epidemic process (proportions of susceptible and infectious (s_0, i_0)) can raise some difficulties when no constraint (such as $s_0 + i_0 = 1$) is imposed and when only one discretized and perturbed coordinate of the system (here I) is observed. This occurred even for a convenient scenario, when $N = 10000$, $n = 100$ and $p = 0.8$ (low stochasticity and little loss of information in the data). The problem is no longer encountered if the two coordinates of the system (S and I) are observed. Besides, two blocks of dependance between the estimates were noticed: (λ, γ, s_0) on the one hand and (p, i_0) on the other hand. Therefore, a wrong estimation of i_0 or s_0 will be reflected in the estimation of p and of (λ, γ), respectively. One potential way to solve this problem could be to consider a prior for the initial conditions of the system. For more details on how to overcome this practical issue, see Stocks et al. (2018), Stocks (2017), who also emphasize the fact that inference algorithms

are very sensitive to the initial values of the system.
Our method relies on two successive Gaussian model approximations (one for the latent state and the other for the observation model). These approximations do not seem to alter the quality of the estimates. Indeed, the small variance coefficient $N^{-1/2}$ provides an advantageous framework for the approximation of the state model for which the Kalman filter has very good performances in practice (small prediction errors). The good accuracy of Gaussian process approximations for stochastic epidemic models was already shown (Buckingham-Jeffery et al. (2018)). Here, we went beyond the investigations in Buckingham-Jeffery et al. (2018) and explored further the performances of Gaussian approximations of epidemic dynamics not only by using a different approach based on Kalman filtering, but also by considering an even less convenient configuration, where the initial condition and the observation error have to be estimated.
Our approach can be generalized in several directions. First, although we focused in this study on the SIR model as a case study, our method is quite generic, since it can be extended to other mechanistic models of epidemic dynamics, including additional health states (such as exposed state E). Second, the observations can encompass variable sampling intervals (i.e. Δ, the time step between two consecutive observations, is not necessarily constant). Third, other types of observations can be considered, both with regard to their nature (e.g. such as the number of new infectious individuals, which can be viewed as a function of state variables S and I) and to the error model.
Hence, given its facility of implementation, low computation time and satisfying performances, we recommend to use our estimation method based on Kalman filtering for providing first guess for parameters, in the framework of partially observed complex epidemic dynamics.

Acknowledgments

We thank two anonymous referees for that constructive and helpful comments.

Funding

This work was supported by the French Agence National de la Recherche [project CADENCE, ANR-16-CE32-0007-01] and by a grant from Région Île-de-France (DIM MathInnov).

Appendix A. Remarks on the sampling interval

The sampling interval Δ is important in our method and we can distinguish two cases: "Small Δ" and "Moderate Δ". We explicit the dependence on the quantities of interest with respect to Δ.
(1) Small sampling interval Δ
The Taylor expansions with respect to t at point t_{k-1} yield

$$\begin{aligned} F_k(\eta) &= F_k(\eta, \Delta) = \Delta \left(b(\eta, x(\eta, t_{k-1})) - \nabla_x b(\eta, x(\eta, t_{k-1}))x(\eta, t_{k-1})\right) + \Delta\, o(1), \\ A_k(\eta) &= A_k(\eta, \Delta) = I_d + \Delta \nabla_x b(\eta, x(\eta, t_{k-1})) + \Delta\, o(1), \\ T_k(\eta) &= T_k(\eta, \Delta) = \frac{1}{N}\left(\Delta \Sigma(\eta, x(\eta, t_{k-1})) + \Delta\, o(1)\right). \end{aligned}$$

These additional approximations that simplify the analytic expressions can be used in the state space equation:

$$X_k = \Delta\left(b(\eta, x(\eta, t_{k-1})) - \nabla_x b(\eta, x(\eta, t_{k-1}))x(\eta, t_{k-1})\right) + \left(I_d + \Delta\nabla_x b(\eta, x(\eta, t_{k-1}))\right) X_{k-1} + U_k,$$
$$U_k \sim \mathcal{N}_d\left(0, \frac{\Delta}{N}\Sigma(\eta, x(\eta, t_{k-1}))\right).$$

(2) Moderate Δ
We can notice that computing the approximated log-likelihood (19) with Kalman filtering techniques requires to compute the resolvent matrix Φ of the ODE system (8). When the time intervals between observations are too large (i.e. Δ too big), we use the approximation for matrix exponential

$$\Phi(\theta_x, t_{k+1}, t_k) \approx \prod_{j=1,\ldots,J-1} \left(I_d + (a_{j+1} - a_j)\nabla_x b(\theta_x, x(\theta_x, a_j))\right), \tag{A.1}$$

where $t_k = a_1 < a_2 < \ldots < a_J = t_{k+1}$. This can however significantly increase computation times.

Appendix B. Proof of Proposition 2

By the semi group property of Φ, we have that g defined in (7) satisfies, for $s \leq t$,

$$\begin{aligned} g(t) &= \Phi(t, s)\int_0^s \Phi(s, u)\sigma(x(u))dB(u) + \int_s^t \Phi(t, u)\sigma(x(u))dB(u), \\ &= \Phi(t, s)g(s) + \int_s^t \Phi(t, u)\sigma(x(u))dB(u). \end{aligned}$$

Substituting $g(s)$ by $\sqrt{N}(G_N(s) - x(s))$ using (9) yields

$$G_N(t) = x(t) + \Phi(t, s)(G_N(s) - x(s)) + \frac{1}{\sqrt{N}}\int_s^t \Phi(t, u)\sigma(x(u))dB(u).$$

Setting $F(t_k) = x(t_k) - \Phi(t_k, t_{k-1})x(t_{k-1})$ and $U_k = \int_{t_{k-1}}^{t_k} \Phi(t_k, u)\sigma(x(u))dB(u)$ yields (ii).
Clearly, U_k is $\mathcal{F}_{t_k}$- measurable. By the property of independent increments of the Brownian motion, we get moreover that U_k is independent of $\mathcal{F}_{t_{k-1}}$. This achieves the proof of Proposition 2.

Appendix C. Proof of Lemma 2

Assume first that Q and T are non singular. The joint distribution of (Y, X) is Gaussian,
$\mathcal{L}(Y, X) \simeq \exp\{-\frac{1}{2}\left((y - Bx)^t Q^{-1}(y - Bx) + (x - \xi)^t T^{-1}(x - \xi)\right)\}$.
Hence, $\mathcal{L}(X|Y) \simeq \exp\{-\frac{1}{2}\left(x^t(B^t Q^{-1}B + T^{-1})x - 2x^t(B^t Q^{-1}y + T^{-1}\xi)\right)\}$.
Setting $\bar{T} = (B^t Q^{-1}B + T^{-1})^{-1} = (I_d + TB^t Q^{-1}B)^{-1}T$, we get
$\mathcal{L}(X|Y) \simeq \exp\{-\frac{1}{2}\left((x - \bar{T}(B^t Q^{-1}y + T^{-1}\xi))^t \bar{T}^{-1}(x - \bar{T}(B^t Q^{-1}y + T^{-1}\xi))\right)\}$,
and $\bar{\xi}(y) = (I_d + TB^t Q^{-1}B)^{-1}T(T^{-1}\xi + B^t Q^{-1}y) = (I_d + TB^t Q^{-1}B)^{-1}(\xi + TB^t Q^{-1}y)$.

We get, using now the following matrix relation,

$$(I_d + TB^tQ^{-1}B)^{-1} = I_d - TB^t(BTB^t + Q)^{-1}B.$$

$$\begin{aligned}
\bar{\xi}(y) &= \xi - TB^t(BTB^t + Q)^{-1}B\xi + TB^t(Q^{-1} - (BTB^t + Q)^{-1}TB^tQ^{-1})y, \\
&= \xi + TB^t(BTB^t + Q)^{-1}(y - B\xi), \\
\bar{T} &= (I_d + TB^tQ^{-1}B)^{-1}T = T - TB^t(BTB^t + Q)^{-1}BT.
\end{aligned}$$

Appendix D. Proof of Proposition 3

For $k = 0$, we have that $X_0 \sim \mathcal{N}(\xi_0, \hat{\Xi}_0)$.The induction assumption is: $\mathcal{L}(X_k|Y_{k-1,0}) = \mathcal{N}_d(\hat{X}_k, \hat{\Xi}_k)$ with $k \geq 1$.
To get (i), we apply Lemma 2 noting that the distribution $\mathcal{L}(X_k|Y_{k-1,0}) = \mathcal{N}_d(\hat{X}_k, \hat{\Xi}_k)$ and that the distribution Y_k conditionally on X_k is $\mathcal{N}(BX_k, Q_k)$. Therefore, setting $\xi = \hat{X}_k$, $T = \hat{\Xi}_k$, B and $Q = Q_k$, we get that the distribution of $(X_k|Y_{k:0})$ is $\mathcal{N}_d(\bar{X}_k, \bar{T}_k)$ with $\bar{X}_k = \bar{\xi}(Y_k)$ where $\bar{\xi}(Y_k)$ and $\bar{T}_k$ are given by (20). It is precisely the expressions for $\bar{T}_k, \bar{\Xi}_k$ given in (i).
For (ii), we use that $X_{k+1} = F_{k+1} + A_kX_k + U_{k+1}$ and $\mathcal{L}(X_k|Y_{k:0}) \sim \mathcal{N}_d(\bar{X}_k, \bar{T}_k)$. Therefore, $\mathcal{L}(X_{k+1}|Y_{k:0}) = \mathcal{N}_d(F_{k+1} + A_k\bar{X}_k, A_k\bar{T}_kA_k^t + T_{k+1})$. Setting $\hat{X}_{k+1} = F_{k+1} + A_k\bar{X}_k$ and $\hat{\Xi}_{k+1} = A_k\bar{T}_kA_k^t + T_{k+1}$ yields (ii).
For (iii), we use that $Y_{k+1} = BX_{k+1} + V_{k+1}$ and that $\mathcal{L}(X_{k+1}|Y_{k:0}) \sim \mathcal{N}(\hat{X}_{k+1}, \hat{\Xi}_{k+1})$. This yields $\mathcal{L}(Y_{k+1}|Y_{k:0})$ is $\mathcal{N}_q(B\hat{X}_{k+1}, B\hat{\Xi}_{k+1}B^t + Q_{k+1})$.
Setting $\hat{M}_{k+1} = B\hat{X}_{k+1}, \hat{\Omega}_{k+1} = B\hat{\Xi}_{k+1}B^t + Q_{k+1}$ yields (iii). The induction assumption is fulfilled and therefore this achieves the proof of Proposition 3.

Appendix E. Complementary simulation study

Appendix E.1. Description

We reproduced the simulation study described in Section 4 with other parameter values: $\lambda = 0.6$, $\gamma = 0.4$, $s_0 = 0.99$, $i_0 = 0.01$. An extract of the simulated data is shown in Figure E.6.

Appendix E.2. Point estimates and standard deviations for key model parameters θ

Appendix E.2.1. Numerical results for the first experiment ($\tau = 0$)

Tables E.6 and E.7 respectively display the results for the high reporting scenario ($p = 0.8$) and the low reporting scenario ($p = 0.3$) when $\tau = 0$ and not estimated. Each Table compares the Kalman-based method (KM) to the Maximum Iterated Filtering algorithm (MIF). The first column display the true parameter values. Columns 2 to 10 display the results for the different combinations (N, n). For each parameter and each estimation method, the reported values are the mean of the 500 parameter estimates and their standard deviations (in brackets).

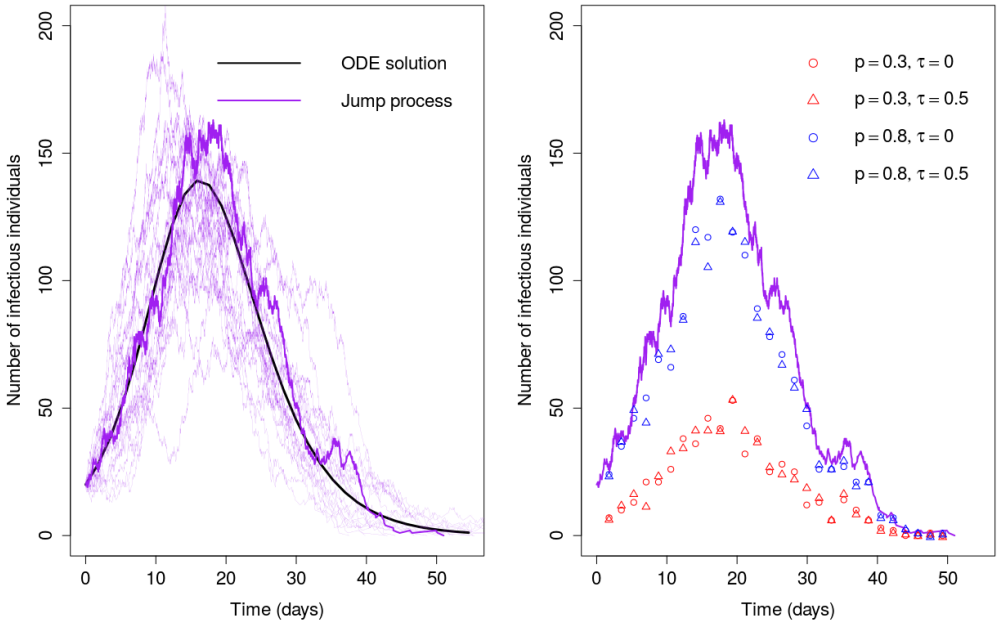

Figure E.6: Left panel : ODE solution for the number of infected individuals I (black plain line) and 20 trajectories of the Markov jump process for I (purple lines) when $N = 2000$. Right panel : $n = 30$ observations obtained from a particular trajectory of the jump process (in purple bold in the left panel), as a function of time. The points and triangles stand for observations generated with measurement error terms $\tau = 0$ and $\tau = 0.5$ respectively and the blue and red symbols represent observations generated with $p = 0.8$ and $p = 0.3$ respectively.

Table E.6: First experiment ($\tau = 0$). Estimation of $\theta = (\lambda, \gamma, p, i_0)$ under the constraint $s_0 + i_0 = 1$ in setting 1 with true parameter values $(\lambda^*, \gamma^*, p^*, i_0^*)$=(0.6, 0.4, 0.8, 0.01). For each combination of (N, n) and for each model parameter, point estimates and standard deviations are calculated as the mean of the 500 individual estimates and their standard deviations (in brackets) obtained by our Kalman-based method (KM) and Maximum Iterated Filtering (MIF). The reported values for the number of observations n correspond to the average over the 500 trajectories, with the min and the max in brackets.

	$N = 1000$			$N = 2000$			$N = 10000$		
	$n = 10$	$n = 30$	$n = 99$	$n = 11$	$n = 31$	$n = 102$	$n = 11$	$n = 31$	$n = 101$
	(3, 19)	(9, 56)	(30, 182)	(7, 19)	(20, 56)	(66, 182)	(8, 17)	(24, 49)	(78, 160)
$\lambda^* = 0.6$									
KM	0.47	0.50	0.59	0.45	0.50	0.59	0.48	0.51	0.60
	(0.16)	(0.15)	(0.16)	(0.08)	(0.10)	(0.08)	(0.04)	(0.06)	(0.05)
MIF	0.50	0.53	0.58	0.49	0.52	0.59	0.51	0.52	0.60
	(0.14)	(0.17)	(0.11)	(0.09)	(0.09)	(0.07)	(0.06)	(0.06)	(0.04)
$\gamma^* = 0.4$									
KM	0.19	0.27	0.39	0.21	0.28	0.39	0.25	0.29	0.40
	(0.08)	(0.11)	(0.09)	(0.07)	(0.09)	(0.04)	(0.05)	(0.07)	(0.03)
MIF	0.22	0.30	0.39	0.25	0.31	0.40	0.29	0.32	0.41
	(0.11)	(0.10)	(0.06)	(0.10)	(0.08)	(0.04)	(0.07)	(0.07)	(0.03)
$p^* = 0.8$									
KM	0.28	0.49	0.75	0.32	0.50	0.77	0.41	0.50	0.82
	(0.20)	(0.27)	(0.11)	(0.18)	(0.22)	(0.08)	(0.13)	(0.18)	(0.09)
MIF	0.37	0.53	0.78	0.40	0.55	0.78	0.49	0.55	0.83
	(0.24)	(0.24)	(0.07)	(0.22)	(0.21)	(0.07)	(0.17)	(0.18)	(0.07)
$i_0^* = 0.01$									
KM	0.029	0.032	0.013	0.025	0.022	0.012	0.018	0.019	0.011
	(0.034)	(0.081)	(0.019)	(0.021)	(0.013)	(0.005)	(0.006)	(0.006)	(0.003)
MIF	0.027	0.025	0.012	0.023	0.019	0.011	0.016	0.016	0.010
	(0.029)	(0.048)	(0.003)	(0.019)	(0.011)	(0.002)	(0.006)	(0.006)	(0.001)

Table E.7: First experiment ($\tau = 0$). Estimation of $\theta = (\lambda, \gamma, p, i_0)$ under the constraint $s_0 + i_0 = 1$ in setting 2 with true parameter values $(\lambda^*, \gamma^*, p^*, i_0^*)$=(0.6, 0.4, 0.3, 0.01). For each combination of (N, n) and for each model parameter, point estimates and standard deviations are calculated as the mean of the 500 individual estimates and their standard deviations (in brackets) obtained by our Kalman-based method (KM) and Maximum Iterated Filtering (MIF). The reported values for the number of observations n correspond to the average over the 500 trajectories, with the min and the max in brackets.

	$N = 1000$			$N = 2000$			$N = 10000$		
	$n = 10$	$n = 30$	$n = 99$	$n = 11$	$n = 31$	$n = 102$	$n = 11$	$n = 31$	$n = 101$
	(3, 19)	(9, 56)	(30, 182)	(7, 19)	(20, 56)	(66, 182)	(8, 17)	(24, 49)	(78, 160)
$\lambda^* = 0.6$									
KM	0.44	0.50	0.53	0.43	0.47	0.54	0.48	0.50	0.55
	(0.18)	(0.12)	(0.15)	(0.08)	(0.09)	(0.09)	(0.06)	(0.06)	(0.07)
MIF	0.47	0.51	0.55	0.47	0.49	0.53	0.52	0.53	0.55
	(0.12)	(0.11)	(0.15)	(0.09)	(0.08)	(0.08)	(0.09)	(0.06)	(0.06)
$\gamma^* = 0.4$									
KM	0.17	0.19	0.29	0.17	0.21	0.31	0.26	0.28	0.34
	(0.19)	(0.09)	(0.09)	(0.06)	(0.08)	(0.08)	(0.07)	(0.06)	(0.08)
MIF	0.20	0.21	0.29	0.22	0.24	0.31	0.31	0.31	0.35
	(0.09)	(0.10)	(0.10)	(0.09)	(0.09)	(0.08)	(0.11)	(0.07)	(0.07)
$p^* = 0.3$									
KM	0.08	0.11	0.19	0.08	0.12	0.21	0.16	0.17	0.24
	(0.08)	(0.08)	(0.09)	(0.04)	(0.07)	(0.09)	(0.07)	(0.07)	(0.09)
MIF	0.11	0.12	0.18	0.12	0.13	0.20	0.21	0.20	0.24
	(0.10)	(0.09)	(0.08)	(0.09)	(0.07)	(0.08)	(0.12)	(0.07)	(0.07)
$i_0^* = 0.01$									
KM	0.028	0.020	0.023	0.023	0.019	0.015	0.020	0.017	0.013
	(0.069)	(0.022)	(0.078)	(0.016)	(0.012)	(0.010)	(0.009)	(0.006)	(0.006)
MIF	0.025	0.022	0.023	0.022	0.020	0.015	0.018	0.015	0.013
	(0.027)	(0.029)	(0.061)	(0.016)	(0.015)	(0.009)	(0.009)	(0.005)	(0.005)

The results on the second set of epidemic parameters are more contrasted, since the parameter values chosen ($\lambda^* = 0.6$ and $\gamma^* = 0.4$) generate more stochasticity (see Figure E.6), so trajectories are less similar and far to the mean of the jump process and hence the estimates are less accurate. Besides, the peak of the number of infectious individuals is clearly lower than in the case $\lambda^* = 1$ and $\gamma^* = 1/3$. The estimations of p are particularly bad when n is low, impacting obviously the estimations of the other parameters.

Appendix E.2.2. Numerical results for the second experiment ($\tau \neq 0$)

Unknown starting point i_0. Table E.8 displays the results obtained by our Kalman-based method and the MIF algorithm for the high reporting scenario ($p = 0.8$).

Table E.8: Second experiment ($\tau \neq 0$). Estimation of $\theta = (\lambda, \gamma, p, i_0, \tau)$ under the constraint $s_0 + i_0 = 1$ in setting 1 with true parameter values $(\lambda^*, \gamma^*, p^*, i_0^*, \tau^*)$=(0.6, 0.4, 0.8, 0.01, 0.5). For each combination of (N, n) and for each model parameter, point estimates and standard deviations are calculated as the mean of the 500 individual estimates and their standard deviations (in brackets) obtained by our Kalman-based method and the MIF algorithm. The reported values for the number of observations n correspond to the average over the 500 trajectories, with the min and the max in brackets.

	$N = 1000$					$N = 10000$				
	$n = 10$	$n = 30$	$n = 99$	$n = 499$	$n = 998$	$n = 11$	$n = 31$	$n = 101$	$n = 500$	$n = 1001$
	(3, 19)	(9, 56)	(30, 182)	(152, 916)	(304, 1831)	(8, 17)	(24, 49)	(78, 160)	(385, 789)	(771, 1577)
$\lambda^* = 0.6$										
KM	0.50	0.49	0.56	0.58	0.57	0.48	0.50	0.55	0.57	0.58
	(0.32)	(0.15)	(0.13)	(0.12)	(0.14)	(0.04)	(0.07)	(0.06)	(0.05)	(0.07)
MIF	0.52	0.51	0.56	0.59	0.58	0.52	0.52	0.56	0.58	0.59
	(0.13)	(0.10)	(0.11)	(0.10)	(0.11)	(0.04)	(0.05)	(0.05)	(0.04)	(0.05)
$\gamma^* = 0.4$										
KM	0.20	0.25	0.35	0.39	0.39	0.25	0.28	0.34	0.38	0.39
	(0.33)	(0.13)	(0.09)	(0.07)	(0.09)	(0.04)	(0.07)	(0.06)	(0.03)	(0.04)
MIF	0.25	0.30	0.36	0.39	0.38	0.30	0.32	0.36	0.39	0.40
	(0.07)	(0.07)	(0.08)	(0.07)	(0.07)	(0.05)	(0.05)	(0.05)	(0.04)	(0.04)
$p^* = 0.8$										
KM	0.24	0.42	0.66	0.77	0.78	0.39	0.49	0.65	0.75	0.77
	(0.16)	(0.25)	(0.23)	(0.16)	(0.14)	(0.11)	(0.20)	(0.19)	(0.11)	(0.11)
MIF	0.39	0.50	0.65	0.72	0.72	0.51	0.55	0.68	0.76	0.79
	(0.16)	(0.17)	(0.17)	(0.13)	(0.15)	(0.13)	(0.15)	(0.15)	(0.13)	(0.13)
$i_0^* = 0.01$										
KM	0.029	0.037	0.022	0.016	0.015	0.019	0.017	0.014	0.012	0.011
	(0.048)	(0.086)	(0.078)	(0.039)	(0.010)	(0.006)	(0.007)	(0.004)	(0.003)	(0.004)
MIF	0.014	0.016	0.014	0.012	0.012	0.015	0.015	0.013	0.011	0.011
	(0.005)	(0.005)	(0.004)	(0.003)	(0.003)	(0.004)	(0.004)	(0.003)	(0.002)	(0.002)
$\tau^* = 0.5$										
KM	0.34	0.44	0.52	0.54	0.52	0.11	0.21	0.35	0.42	0.47
	(0.69)	(0.44)	(0.20)	(0.16)	(0.17)	(0.22)	(0.21)	(0.20)	(0.13)	(0.13)
MIF	0.49	0.49	0.47	0.48	0.43	0.46	0.39	0.35	0.44	0.49
	(0.25)	(0.19)	(0.17)	(0.15)	(0.20)	(0.24)	(0.18)	(0.17)	(0.15)	(0.14)

Known starting point i_0. Tables E.9 and E.10 respectively display the results obtained by our Kalman-based method and the MIF algorithm for the high reporting scenario ($p = 0.8$) and the low reporting scenario ($p = 0.3$).

Table E.9: Second experiment ($\tau \neq 0$). Estimation of $\theta = (\lambda, \gamma, p, \tau)$ with $s_0 = 0.99$ and $i_0 = 0.01$ known in setting 1 with true parameter values $(\lambda^*, \gamma^*, p^*, \tau^*)$=$(0.6, 0.4, 0.8, 0.5)$. For each combination of (N, n) and for each model parameter, point estimates and standard deviations are calculated as the mean of the 500 individual estimates and their standard deviations (in brackets) obtained by KM and MIF. The reported values for the number of observations n correspond to the average over the 500 trajectories, with the min and the max in brackets.

	$N = 1000$			$N = 2000$			$N = 10000$		
	$n = 10$	$n = 30$	$n = 99$	$n = 11$	$n = 31$	$n = 102$	$n = 11$	$n = 31$	$n = 101$
	$(3, 19)$	$(9, 56)$	$(30, 182)$	$(7, 19)$	$(20, 56)$	$(66, 182)$	$(8, 17)$	$(24, 49)$	$(78, 160)$
$\lambda^* = 0.6$									
KM	0.54	0.57	0.59	0.55	0.58	0.60	0.56	0.56	0.59
	(0.15)	(0.15)	(0.13)	(0.11)	(0.10)	(0.08)	(0.06)	(0.06)	(0.05)
MIF	0.55	0.55	0.59	0.56	0.57	0.60	0.57	0.58	0.60
	(0.12)	(0.10)	(0.11)	(0.09)	(0.07)	(0.07)	(0.04)	(0.03)	(0.03)
$\gamma^* = 0.4$									
KM	0.27	0.35	0.38	0.30	0.36	0.39	0.35	0.36	0.39
	(0.11)	(0.11)	(0.10)	(0.09)	(0.07)	(0.05)	(0.05)	(0.05)	(0.04)
MIF	0.27	0.34	0.39	0.30	0.37	0.40	0.36	0.39	0.41
	(0.09)	(0.07)	(0.07)	(0.09)	(0.05)	(0.04)	(0.04)	(0.03)	(0.02)
$p^* = 0.8$									
KM	0.42	0.61	0.72	0.49	0.67	0.76	0.65	0.69	0.77
	(0.22)	(0.22)	(0.18)	(0.20)	(0.19)	(0.15)	(0.14)	(0.15)	(0.12)
MIF	0.43	0.62	0.75	0.49	0.67	0.78	0.67	0.75	0.80
	(0.19)	(0.17)	(0.13)	(0.19)	(0.14)	(0.10)	(0.12)	(0.09)	(0.07)
$\tau^* = 0.5$									
KM	0.41	0.63	0.60	0.32	0.65	0.62	0.09	0.30	0.46
	(0.47)	(0.33)	(0.20)	(0.42)	(0.25)	(0.16)	(0.26)	(0.28)	(0.16)
MIF	0.52	0.56	0.55	0.55	0.62	0.56	0.50	0.50	0.47
	(0.27)	(0.20)	(0.13)	(0.29)	(0.18)	(0.12)	(0.27)	(0.19)	(0.10)

Table E.10: Second experiment ($\tau \neq 0$). Estimation of $\theta = (\lambda, \gamma, p, \tau)$ with $s_0 = 0.99$ and $i_0 = 0.01$ known in setting 2 with true parameter values $(\lambda^*, \gamma^*, p^*, \tau^*)$=(0.6, 0.4, 0.3, 0.5). For each combination of (N, n) and for each model parameter, point estimates and standard deviations are calculated as the mean of the 500 individual estimates and their standard deviations (in brackets) obtained by KM and MIF. The reported values for the number of observations n correspond to the average over the 500 trajectories, with the min and the max in brackets.

	$N = 1000$			$N = 2000$			$N = 10000$		
	$n = 10$	$n = 30$	$n = 99$	$n = 11$	$n = 31$	$n = 102$	$n = 11$	$n = 31$	$n = 101$
	(3, 19)	(9, 56)	(30, 182)	(7, 19)	(20, 56)	(66, 182)	(8, 17)	(24, 49)	(78, 160)
$\lambda^* = 0.6$									
KM	0.47	0.57	0.57	0.47	0.54	0.58	0.56	0.55	0.56
	(0.16)	(0.13)	(0.16)	(0.12)	(0.09)	(0.09)	(0.06)	(0.06)	(0.05)
MIF	0.51	0.55	0.55	0.52	0.55	0.56	0.57	0.57	0.59
	(0.12)	(0.13)	(0.11)	(0.09)	(0.07)	(0.07)	(0.05)	(0.04)	(0.03)
$\gamma^* = 0.4$									
KM	0.20	0.29	0.35	0.26	0.30	0.37	0.37	0.35	0.35
	(0.12)	(0.13)	(0.14)	(0.10)	(0.09)	(0.09)	(0.05)	(0.06)	(0.05)
MIF	0.22	0.26	0.32	0.25	0.30	0.34	0.38	0.37	0.39
	(0.08)	(0.10)	(0.09)	(0.10)	(0.08)	(0.06)	(0.06)	(0.04)	(0.03)
$p^* = 0.3$									
KM	0.10	0.19	0.24	0.15	0.20	0.27	0.28	0.25	0.25
	(0.09)	(0.15)	(0.12)	(0.08)	(0.12)	(0.13)	(0.06)	(0.08)	(0.10)
MIF	0.11	0.14	0.19	0.14	0.18	0.21	0.27	0.27	0.28
	(0.06)	(0.09)	(0.06)	(0.08)	(0.07)	(0.05)	(0.07)	(0.04)	(0.03)
$\tau^* = 0.5$									
KM	0.23	0.42	0.52	0.13	0.24	0.51	0.05	0.19	0.36
	(0.20)	(0.29)	(0.21)	(0.19)	(0.21)	(0.18)	(0.14)	(0.15)	(0.12)
MIF	0.38	0.37	0.44	0.39	0.29	0.43	0.61	0.33	0.43
	(0.14)	(0.15)	(0.09)	(0.15)	(0.15)	(0.07)	(0.17)	(0.12)	(0.05)

Appendix E.3. Numerical confidence intervals

Figures E.7, E.8 represent the profile likelihoods and the subsequent confidence intervals (CI95%) for the parameters λ and γ obtained for our Kalman filtering based method on 2 cases (first case: $N = 2000$, $n = 30$ and $p = 0.3$; second case: $N = 10000$, $n = 100$ and $p = 0.8$).

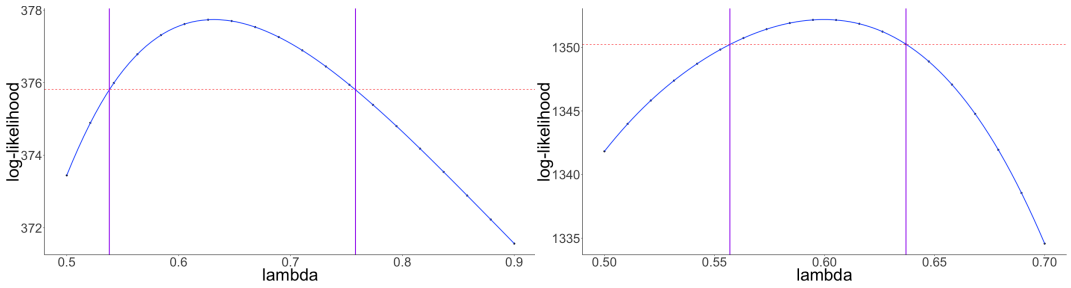

Figure E.7: Profile likelihood and confidence intervals (CI95%) for λ. Left panel: $N = 2000$, $n = 30$ and $p = 0.3$. The true value $\lambda^* = 0.6$, the point estimate $\hat{\lambda} = 0.47$ and CI95% = [0.54, 0.76]. Right panel : $N = 10000$, $n = 100$ and $p = 0.8$. The true value $\lambda^* = 0.6$, the point estimate $\hat{\lambda} = 0.60$ and CI95% = [0.56, 0.64].

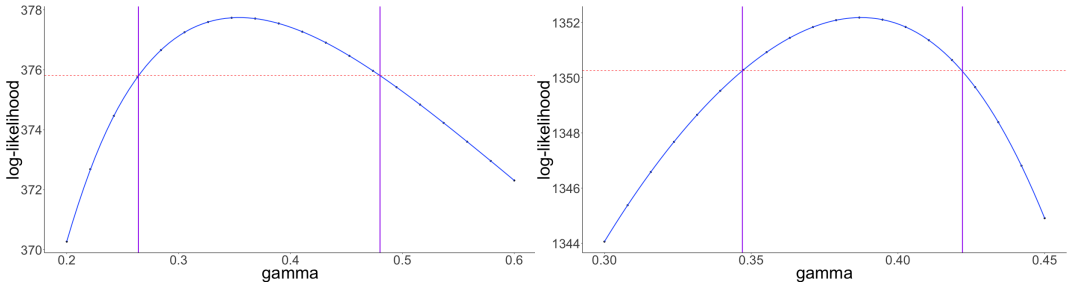

Figure E.8: Profile likelihood and confidence intervals (CI95%) for γ. Left panel: $N = 2000$, $n = 30$ and $p = 0.3$. The true value $\gamma^* = 0.4$, the point estimate $\hat{\gamma} = 0.21$ and CI95% $= [0.26, 0.48]$. Right panel : $N = 10000$, $n = 100$ and $p = 0.8$. The true value $\gamma^* = 0.4$, the point estimate $\hat{\gamma} = 0.40$ and CI95% $= [0.35, 0.42]$.

Appendix F. User-friendly code

We proposed a user-friendly code composed of four distinct programs in R language and available on the RunMyCode website at the following link: `http://www.runmycode.org/companion/view/4074`.

KalmanFunctions.R includes generic functions implementing the Kalman filter and computing the likelihood of the observations, given a specified compartmental model, with a fixed sampling interval. These functions are easily generalizable to the case where the sampling interval is variable. Moreover, this script includes a function computing the resolvent matrix for high sampling interval.

ModelFunctions.R implements SIR and SEIR models and defines the key quantities (described in the manuscript for the SIR model) necessary to apply the Kalman filter based method. More precisely, given a compartmental model (SIR or SEIR), the following functions are implemented: the ode system, the drift function, the gradient of the drift function, the diffusion matrix, the projection operator linking the observations to the states of the epidemic model and the variance of the observations.

SIRexample.R and *SEIRexample.R* simulate respectively SIR and SEIR Markovian jump processes for a set of parameters values by using the package `GillespieSSA`. The observations of infectious individuals are obtained as $O_1(t_k) = \mathcal{B}(I(t_k), p)$, $O_2(t_k) = \mathcal{N}(0, \tau^2 I(t_k))$, $k = 1, ..., n$, at regularly space time points. Finally, an estimation of key parameters λ, γ, p and τ with known starting points and, in the SEIR model, with a known transition rate from E to I, is proposed.

References

Andersson, H., Britton, T., 2000. Stochastic epidemic models and their statistical analysis. volume 151 of *Lecture Notes in Statistics*. Springer. doi:10.1007/978-1-4612-1158-7.

Andrieu, C., Doucet, A., Holenstein, R., 2010. Particle markov chain monte carlo methods. Journal of the Royal Statistical Society B 72, 269–342. doi:10.1111/j.1467-9868.2009.00736.x.

Anonymous, 1978. Influenza in a boarding school. British Medical Journal .

Azencott, R., 1982. Formule de taylor stochastique et développement asymptotique intégrales de feynmann. Séminaire de Probabilités XVI , 237–285.URL: http://www.numdam.org/item/SPS_1982__S16__237_0.

Britton, T., Giardina, F., 2016. Introduction to statistical inference for infectious diseases. Journal de la Société Française de Statistique 157, 53–70. arXiv:1411.3138.

Britton, T., Pardoux, E., 2020. Stochastic epidemic models with inference. Springer. doi:10.1007/978-3-030-30900-8.

Buckingham-Jeffery, E., Isham, V., House, T., 2018. Gaussian process approximations for fast inference from infectious disease data. Mathematical Biosciences 301, 111 – 120. doi:10.1016/j.mbs.2018.02.003.

Cauchemez, S., Ferguson, N.M., 2008. Likelihood-based estimation of continuous-time epidemic models from time-series data: application to measles transmission in london. Journal of The Royal Society Interface 5, 885–897. doi:10.1098/rsif.2007.1292.

Ethier, S.N., Kurtz, T.G., 2005. Markov processes: characterization and convergence. Wiley, 2nd edition. doi:10.1002/9780470316658.

Favetto, B., Samson, A., 2010. Parameter estimation for a bidimensional partially observed ornstein-uhlenbeck process with biological application. Scandinavian Journal of Statistics 37, 200–220. doi:10.1111/j.1467-9469.2009.00679.x.

Freidlin, M., Wentzell, A., 1978. Random perturbations of dynamical systems. Springer doi:10.1007/978-3-642-25847-3.

Gillespie, D.T., 1977. Exact stochastic simulation of coupled chemical reactions. The Journal of Physical Chemistry 81, 2340–2361. doi:10.1021/j100540a008.

Guy, R., Larédo, C., Vergu, E., 2014. Parametric inference for discretely observed multidimensional diffusions with small diffusion coefficient. Stochastic Processes and their Applications 124, 51–80. doi:10.1016/j.spa.2013.07.009.

Guy, R., Larédo, C., Vergu, E., 2015. Approximation of epidemic models by diffusion processes and their statistical inference. J. Math. Bio 70, 621–646. doi:10.1007/s00285-014-0777-8.

Ionides, E.L., Bhadra, A., Atchadé, Y., King, A.A., 2011. Iterated filtering. The Annals of Statistics 39, 1776–1802. doi:10.1214/11-aos886.

Ionides, E.L., Breto, C., King, A.A., 2006. Inference for nonlinear dynamical systems. Proceedings of the National Academy of Sciences of the United States of America 103, 18438–18443. doi:10.1073/pnas.0603181103.

Ionides, E.L., Breto, C., Park, J., Smith, R.A., King, A.A., 2017. Monte carlo profile confidence intervals for dynamic systems. J. R. Soc Interface 14, 2017126. doi:10.1098/rsif.2017.0126.

Ionides, E.L., Nguyen, D., Atchadé, Y., Stoev, S., King, A.A., 2015. Inference for dynamic and latent variable models via iterated, perturbed bayes maps. Proceedings of the National Academy of Sciences of the United States of America 112, 719–724. doi:10.1073/pnas.1410597112.

King, A.A., Nguyen, D., Ionides, E.L., 2017. Statistical inference for partially observed markov processes via the r package pomp. Journal of Statistical Software 69, 1–43. doi:10.18637/jss.v069.i12.

Norris, J.R., 1997. Markov chains. Cambridge University Press. doi:10.1017/CBO9780511810633.

O'Neill, P.D., 2010. Introduction and snapshot review: Relating infectious disease transmission models to data. Statistics in Medicine 29, 2069–2077. doi:10.1002/sim.3968.

Sisson, S., Fan, Y., Tanaka, M., 2007. Sequential monte carlo without likelihoods. Proceedings of the National Academy of Sciences of the United States of America 104, 1760–1765. doi:10.1073/pnas.0607208104.

Stocks, T., 2017. Iterated filtering methods for markov process epidemic models. arXiv:1712.03058.

Stocks, T., Britton, T., Höhle, M., 2018. Model selection and parameter estimation for dynamic epidemic models via iterated filtering: application to rotavirus in germany. Biostatistics 21, 400–416. doi:10.1093/biostatistics/kxy057.

Toni, T., Welch, D., Strelkowa, N., Ipsen, A., P.H Stumpf, M., 2009. Approximate bayesian computation scheme for parameter inference and model selection in dynamical systems. Journal of The Royal Society Interface 6, 187–202. doi:10.1098/rsif.2008.0172.