Journal de la Société Française de Statistique
Vol. 157 No. 1 (2016)

Approximation and inference of epidemic dynamics by diffusion processes

Titre: Approximation et inférence des dynamiques épidémiques par des processus de diffusion

Romain Guy[1,2], Catherine Larédo[1,2] and Elisabeta Vergu[1]

Abstract: Epidemic data are often aggregated and partially observed. Parametric inference through likelihood-based approaches is rarely straightforward, whatever the mathematical representation used. Recent data augmentation and likelihood-free methods do not completely circumvent the issues related to incomplete data in practice, mainly due to the size of missing data and to the various tuning parameters to be adjusted. In this context, diffusion processes provide a good approximation of epidemic dynamics and allow shedding new light on inference problems related to epidemic data. In this article we summarize and extend previous work on the elaboration of a statistical framework to deal with epidemic models and epidemic data using multidimensional diffusion processes with small diffusion coefficient. First, we construct multidimensional diffusion processes with small variance as mathematical representations of epidemic dynamics, by approximating Markov jump processes. Second, we introduce an inference method related to the asymptotic of the small diffusion coefficient on a fixed time interval for the parameters of the diffusion processes obtained, when all the coordinates are discretely observed. Consistency and asymptotic normality of estimators for this case are obtained for parameters in drift (high and low frequency observations) and diffusion (high frequency observations) coefficients. Third, as an extension of previous work, the case of incomplete data, when only one coordinate of the system is observed, is considered for high frequency observations. Finally, the performances of our estimators are explored for single outbreaks (*SIR* model, simulated data) and for recurrent outbreaks (*SIRS* model, simulated and observed data).

Résumé : Les données épidémiques sont souvent agrégées et partiellement observées. L'inférence paramétrique par des approches de vraisemblance est rarement pratiquable, indépendamment du formalisme mathématique utilisé. Les méthodes récentes d'augmentation de données ou sans vraisemblance ne permettent pas de résoudre définitivement le problème des données incomplètes en pratique, notamment à cause de la taille des données à compléter et des différents paramètres algorithmiques d'ajustement. Dans ce contexte, les processus de diffusion fournissent de bonnes approximations des dynamiques épidémiques et apportent un nouvel éclairage aux problèmes d'inférence liés aux données épidémiques. Dans cet article, nous résumons et complétons des travaux précédents sur l'élaboration d'un cadre statistique pour traiter les modèles et les données épidémiques en utilisant des processus de diffusion multidimensionnels à petite variance. Premièrement, nous construisons de tels processus, comme représentations mathématiques des dynamiques épidémiques, en approximant des processus Markoviens de sauts. Deuxièmement, nous introduisons une méthode d'inférence dans le cadre asymptotique de la petite variance sur un intervalle de temps fixe pour les paramètres des processus de diffusion obtenus, lorsque toutes les coordonnées du système sont observées de façon discrétisée. La convergence et la normalité asymptotique des estimateurs sont obtenues dans ce cas pour les paramètres de la dérive (observations haute et basse fréquence) et du coefficient de diffusion (observations haute fréquence). Troisièmement, en prolongation de travaux précédents, nous étudions le cas des données incomplètes, lorsque seulement une coordonnée du système est observée, pour des observations haute fréquence. Finalement, les performances de nos estimateurs sont explorées pour une seule vague épidémique (modèle *SIR*, données simulées) et pour des épidémies récurrentes (modèle *SIRS*, données simulées et observées).

[1] MaIAGE, INRA, Université Paris-Saclay, 78350 Jouy-en-Josas, France.
[2] UMR 7599 Laboratoire de Probabilités et Modèles Aléatoires, Université Denis Diderot Paris 7 and CNRS, Paris, France.
E-mail: catherine.laredo@jouy.inra.fr and E-mail: elisabeta.vergu@jouy.inra.fr

 ISSN: 2102-6238

Keywords: epidemic data, parametric inference, discrete observations, partially observed processes.
Mots-clés : donnés épidémiques, inférence paramétrique, observations discrétisées, processus partiellement observés.
AMS 2000 subject classifications: 6F12, 60J27, 60J60

1. Introduction

The increasing role of mathematical models in understanding dynamics of pathogen spread at various temporal and spatial scales, their prevention and control is now well recognized. These models are by essence mechanistic and naturally include parameters in their description. The insights provided by models are even more valuable when their parameters are estimated from data.

Typically, as summarized in the article by Britton and Giardina (2014) in this journal issue, epidemic dynamics in closed population of size N can be described by *SIR* models based on three health states (Susceptible, Infectious and Removed from the infectious chain, most usually following immunisation), which form a partition of the population. The *SIR* model can be naturally formalized by a two-dimensional continuous-time Markov jump process. Since the population is closed, for all t, we have $R(t) = N - S(t) - I(t)$ and we can define the process $Z(t) = (S(t), I(t))$ characterized by the initial state $Z(0) = (S(0), I(0))$ and transitions $(S,I) \longrightarrow (S-1, I+1)$ and $(S,I) \longrightarrow (S, I-1)$. The *SIRS* variant of this model assumes an additional transition $(S,I) \longrightarrow (S+1, I)$ accounting for the return at the susceptible state from the R state.

When data are available, key parameters can be estimated using these models through likelihood-based or M-estimation methods sometimes coupled to Bayesian methods (Diekmann et al., 2013). However, these data are most often partially observed (e.g. 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. In this case, estimation through likelihood-based approaches is rarely straightforward, regardless to the mathematical formalism. As summarised in section 5 of Britton and Giardina (2014), various approaches were developed during the last years to deal with partially observed epidemics. Data augmentation and likelihood-free methods (such as the approximate Bayesian computation) opened some of the most promising pathways for improvement (Bretó et al., 2009; McKinley et al., 2009). Nevertheless, these methods do not completely circumvent the issues related to incomplete data. As stated also in Britton and Giardina (2014), there are some limitations in practice, due to the size of missing data and to the various tuning parameters to be adjusted (Andersson and Britton, 2000; O'Neill, 2010). Moreover, identifiability issues are rarely addressed.

In this context, it appears that diffusion processes, satisfactorily approximating epidemic dynamics (see e.g. Fuchs, 2013; Ross et al., 2009), can be profitably used for inference of model parameters from epidemiological data, due to their analytical power. After normalizing the *SIR* Markov jump process $(Z(t))$ by N, an ODE system is asymptotically obtained as the population size N goes to infinity: $x(t) = (s(t), i(t))$ with $(s(0) = 1 - i_0, i(0) = i_0)$.

A first order approximation of the process $(Z(t)/N)_{t\geq 0}$ with respect to N leads to a diffusion process $(X_N(t))_{t\geq 0}$ with diffusion coefficient proportional to $1/\sqrt{N}$. Hence, epidemic dynamics can be described using the two-dimensional diffusion $(X_N(t))_{t\geq 0}$ with a small diffusion coefficient proportional to $1/\sqrt{N}$. Since epidemics are usually observed over limited time periods, we con-

 ISSN: 2102-6238

sider in what follows the parametric inference based on observations of the epidemic dynamics on a fixed interval $[0,T]$.

In this paper, we summarize and complete previous work on the use of diffusion processes with small diffusion coefficient to study epidemic dynamics (Guy et al., 2014, 2015). By extending results stated in (Ethier and Kurtz, 2005, chapter 8), for density-dependent population processes, we develop here a generic and rigorous method to construct multidimensional diffusion processes with small variance as mathematical representations of epidemic dynamics, by approximating a Markov jump process (section 2). Section 3 contains assumptions, notations, statistical framework and a short recap of classical results on the parametric inference for diffusion processes. In section 4, we introduce an inference method for the parameters of the diffusion process obtained in section 2, when the data consist of n discrete observations of $(X(t))$ with sampling interval Δ on a fixed time interval $[0,T]$. Concerning the application of diffusion approximations to epidemic data, we consider the special case where the same parameters are both present in the diffusion and drift terms, the number n of observations satisfying $n\Delta = T$. We consider the case where both coordinates of the diffusion are observed (corresponding here to the observation of susceptible and infected individuals). The consistency and asymptotic normality of estimators are investigated in the two distinct asymptotic frameworks: the number of observations n is fixed (low frequency observations corresponding to $\Delta = T/n$) and the population size $N \to \infty$; $n \to \infty$ (high-frequency observations corresponding to $\Delta = \Delta_n \to 0$) and $N \to \infty$ simultaneously. Motivated by epidemics where only infected individuals can be observed, the novel case of partial data (observation of one coordinate of the system only) is investigated in section 5. Finally, the accuracy of these estimations is explored (section 6) for single outbreaks (*SIR* model, simulated data) and for recurrent outbreaks (*SIRS* model, simulated and observed data).

2. Approximation of jump processes describing epidemic dynamics by diffusion models

This section presents a general approach for building multidimensional diffusion processes with small diffusion coefficient by approximating Markov jump processes, based on Guy et al. (2015). We first detail our approach which is the basis for the statistical inference. Using limit theorems for stochastic processes, we characterize the limiting Gaussian process. Then, based on the theory of perturbations of dynamical systems (Freidlin and Wentzell, 1978), we link the normalised process to a diffusion process with small diffusion coefficient (2.1). These approximations are then applied to *SIR* and *SIRS* models for epidemic dynamics (2.2).

2.1. The approximation scheme

The simplest stochastic model for describing epidemic dynamics in closed populations of size N is a multidimensional Markov jump process $(Z(t), t \geq 0)$ with state space $E \subset \mathbb{Z}^p$, where p corresponds to the number of health states assumed in the model of the infection dynamics. This process is usually described by an initial distribution on E, and a collection of non negative functions $(\alpha_l(.) : E \to \mathbb{R}^+)$ indexed by $l \in \mathbb{Z}^p, l \neq (0,\ldots,0)$ that satisfy,

$$\forall k \in E, 0 < \sum_{l \in \mathbb{Z}^p} \alpha_l(k) = \alpha(k) < \infty. \tag{1}$$

 ISSN: 2102-6238

These functions are the transition rates of the process $(Z(t))$ with Q-matrix having as elements

$$q_{k,k+l} = \alpha_l(k) \text{ if } l \neq 0, \text{ and } q_{k,k} = -\alpha(k) \text{ for } k, k+l \in E. \tag{2}$$

Another useful description of $(Z(t))$ is based on the joint distribution of its jump chain and holding times. The process $Z(t)$ stays in each state $k \in E$ during an exponential time $\mathscr{E}(\alpha(k))$, and then jumps to the state $k+l$ according to a Markov chain (X_n) with transition probabilities $\mathbb{P}(X_{n+1} = k+l \mid X_n = k) = \alpha_l(k)/\alpha(k)$.

An important class of pure jump Markov processes in a closed population consists of the density dependent Markov jump processes which possess a limit behaviour when normalised by the population size N. Let us define the two sets

$$E = \{0, \ldots, N\}^p; \quad E^- = \{-N, \ldots, N\}^p. \tag{3}$$

In the sequel, we assume that the state space of $(Z(t))$ is E and that its jumps belong to E^-. From the original jump process $(Z(t))$ on $E = \{0, \ldots, N\}^p$, one defines a family of normalized jump Markov process $(Z_N(t), t \geq 0)$ where

$$Z_N(t) = \frac{Z(t)}{N} \text{ with state space } E_N = \{N^{-1}k, k \in E\}. \tag{4}$$

Its jumps are now $z = l/N$ with $l \in E^-$ and transition rates from $y \in E_N$ to $y + z \in E_N$ defined using (2),

$$q^N_{y,y+z} = \alpha_{Nz}(Ny) \ \ (= \alpha_l(k) \text{ if } y = k/N, z = l/N). \tag{5}$$

Denote for $y = (y_1, \ldots, y_p) \in \mathbb{R}^p$, $[y] = ([y_1], \ldots, [y_p]) \in \mathbb{Z}^p$, where $[y_i]$ is the integer part of y_i.

We assume in the sequel that there exists a collection of functions $\beta_l : [0,1]^p \to \mathbb{R}^+$ such that,

(H1): $\forall l \in E^-$, $\forall y \in [0,1]^p$ $\ \frac{1}{N}\alpha_l([Ny]) \underset{N \to +\infty}{\longrightarrow} \beta_l(y)$;
(H2): $\forall l \in E^-$, $\beta_l \in C^2([0,1]^p)$.

A jump process $(Z(t))$ with state space E (see (3)) is density dependent if it satisfies (H1). Then, define the two functions b_N and $b : [0,1]^p \to \mathbb{R}^p$ and the two $p \times p$ positive symmetric matrices Σ_N and Σ (with the notation ${}^tM$ for the transposition of a matrix or vector) by

$$b_N(y) = \frac{1}{N} \sum_{l \in E^-} l \, \alpha_l([Ny]), \ \ b(y) = \sum_{l \in E^-} l \, \beta_l(y) \text{ and} \tag{6}$$

$$\Sigma_N(y) = \frac{1}{N} \sum_{l \in E^-} \alpha_l([Ny]) \, l \, {}^t l, \ \ \Sigma(y) = \sum_{l \in E^-} \beta_l(y) \, l \, {}^t l. \tag{7}$$

Since the number of jumps is finite, the function $b(.)$ is well defined under (H1) and Lipschitz under (H2). Therefore, there exists a unique smooth solution $x(t)$ to the ODE

$$x(t) = x_0 + \int_0^t b(x(u))du. \tag{8}$$

 ISSN: 2102-6238

The resolvent matrix $\Phi(t,u)$ associated with (8) is defined as the solution

$$\frac{d\Phi}{dt}(t,u) = \frac{db}{dx}(x(t))\Phi(t,u); \quad \Phi(u,u) = I_p, \text{ where } \frac{db}{dx}(x) = \left(\frac{\partial b_i}{\partial x_j}(x)\right)_{1\le i,j\le p}. \tag{9}$$

Based on notations, definitions and assumptions introduced above, we can run through two distinct approaches to build approximations of the family of processes $(Z_N(t))$. The first approach is similar to (Ethier and Kurtz, 2005, Chapter 9) but proofs are based on different tools. While the convergence results of Ethier and Kurtz (2005) are based on an explicit expression of pure jump Markov processes as sums of random time changed Poisson processes, we use in Guy et al. (2015) general convergence theorems for Markov processes (Jacod and Shiryaev, 1987) to obtain convergence results for $(Z_N(t))$ (almost sure convergence and associated central limit theorem) in the case of processes satisfying (H1), (H2). The second approach we used relies on the theory of small perturbations of dynamical systems (Freidlin and Wentzell, 1978). We obtained that the first order approximation of $(Z_N(t))$ with respect to N could be derived from a diffusion process $(X_N(t))$ with small diffusion coefficient. All these results are detailed below in Theorems 2.1 and 2.2.

Let $(D,\mathscr{D})$ denote the space of "cadlag" functions $f:\mathbb{R}^+\to\mathbb{R}^p$ endowed with the Skorokhod topology $\mathscr{D}$ and denote, for $y\in\mathbb{R}^p$, $\| y \|$ its Euclidean norm. We can now detail the first approach which corresponds to a central limit theorem associated to the convergence of $(Z_N(t))$ to a Gaussian process when $N\to\infty$.

Theorem 2.1. *Assume (H1), (H2) and that $Z_N(0)\to x_0$ as $N\to\infty$. Then, using definitions (4), (7), (8) and (9)*

$$\forall t\ge 0, \lim_{N\to\infty}\sup_{u\le t}\| Z_N(u)-x(u)\| = 0 \ \ a.s. \tag{10}$$

where $x(u)$ is solution of (8).
The process $\sqrt{N}(Z_N(t)-x(t))_{t\ge 0}$ belongs to D and satisfies

$$\sqrt{N}(Z_N(t)-x(t))_{t\ge 0}\to (G(t)) \text{ in distribution in } (D,\mathscr{D}), \tag{11}$$

where $(G(t))$ is a centered Gaussian process with covariance matrix

$$Cov(G(t),G(r)) = \int_0^{t\wedge r}\Phi(t,u)\Sigma(x(u))\,{}^t\Phi(r,u)du. \tag{12}$$

The proofs of these results are given in Guy et al. (2014), Guy et al. (2015).

Note that there is no time scaling here, but only a scaling of the state space. Therefore these results naturally extend to time-dependent Markov jump processes $(Z(t))$ with transition rates $\alpha_l(t,k)$ with $\alpha(t,k)=\sum_{l\in\mathbb{Z}^p}\alpha_l(t,k)<\infty$ under the assumptions,

(H1b): $\forall t\ge 0$, $\forall l\in E^-$, $\forall y\in[0,1]^p$ $\frac{1}{N}\alpha_l(t,[Ny])\underset{N\to+\infty}{\longrightarrow}\beta_l(t,y)$;

(H2b): $\forall l\in E^-$, $\beta_l(t,y)\in C^2(\mathbb{R}^+,[0,1]^p)$.

 ISSN: 2102-6238

The second approach for approximating $(Z_N(t))$ rests on a generator based expansion. Indeed, the process $(Z_N(t))$ in (4) is a Markov process with generator $\mathscr{A}_N$ defined for $f \in C^2(\mathbb{R}^p, \mathbb{R})$ by

$$\mathscr{A}_N f(y) = \sum_{l \in E^-} \alpha_l(Ny)(f(y + \frac{l}{N}) - f(y)).$$

Heuristically, a Taylor expansion of $\mathscr{A}_N f(y)$ yields, using (H1), (H2) and (6),

$$\begin{aligned} \mathscr{A}_N f(y) &= \textstyle\sum_{l \in E^-} N\beta_l(y)(f(y + \frac{l}{N}) - f(y)) + o(1/N) \\ &= b(y)\nabla f(y) + \frac{1}{2N}\left(\textstyle\sum_{l \in E^-} \beta_l(y) \sum_{i,j=1}^p l_i l_j \frac{\partial^2 f}{\partial y_i \partial y_j}(y)\right) + o(1/N) \\ &= b(y)\nabla f(y) + \frac{1}{2N}\textstyle\sum_{i,j=1}^p \Sigma_{ij}(y)\frac{\partial^2 f}{\partial y_i \partial y_j}(y) + o(1/N), \end{aligned}$$

where the last equality is obtained using (7).
The first two terms of the last expression correspond to the generator of a diffusion process on $\mathbb{R}^p$ with drift coefficient $b(.)$ and diffusion matrix $\frac{1}{N}\Sigma(.)$,

$$dX_N(t) = b(X_N(t))dt + \frac{1}{\sqrt{N}}\sigma(X_N(t))dB(t)\ ;\ \ X_N(0) = Z_N(0), \tag{13}$$

where $(B(t)_{t \geq 0})$ is a Brownian motion on $\mathbb{R}^p$ defined on a probability space $\mathbb{P} = (\Omega, (\mathscr{F}_t)_{t \geq 0}, P)$ independent of $Z_N(0)$, and $\sigma(.)$ is a square root of $\Sigma(.)$,

$$\sigma(.) : \mathbb{R}^p \to \mathbb{R}^p \times \mathbb{R}^p \text{ such that } \ \sigma(y)\ {}^t\sigma(y) = \Sigma(y).$$

Let us note that, contrary to the first approach, this does not yield an approximation result between the sample paths of (Z_N) and (X_N). Indeed, no similar limit theorem can justify that (X_N) approximates (Z_N), since the (Z_N) are not expressible in terms of any kind of limiting process (see Ethier & Kurtz , chapter 11, section 3). A coupling theorem of Kolmos-Major-Tusnady is required to get a direct comparison. The main interest of this second approach lies in the fact that it allows all the mathematical developments available for diffusion processes to be used in our framework.

However, these two approaches can be linked a posteriori using the theory of random perturbations of dynamical systems (or stochastic Taylor expansion of diffusion processes) (Freidlin and Wentzell, 1978; Azencott, 1982) and the following theorem.

Theorem 2.2. *Setting $\varepsilon = 1/\sqrt{N}$, the paths of $X_N(.)$ satisfy, as $\varepsilon \to 0$,*

$$X_N(t) = X_\varepsilon(t) = x(t) + \varepsilon g(t) + \varepsilon^2 R_\varepsilon(t), \ \ \textit{with } sup_{t \leq T} \parallel \varepsilon R_\varepsilon(t) \parallel \to 0 \textit{ in probability,} \tag{14}$$

where $x(t)$ is the solution of (8) and $(g(t))$ is the process satisfying the SDE

$$dg(t) = \frac{db}{dx}(x(t))g(t)dt + \sigma(x(t))dB(t), \ \ g(0) = 0.$$

This stochastic differential equation can be solved explicitly and we get a closed form for its solution $g(.)$,

$$g(t) = \int_0^t \Phi(t,s)\sigma(x(s))dB(s). \tag{15}$$

 ISSN: 2102-6238

Hence, $(g(t))$ is a centered Gaussian process having the same covariance matrix (12) as the process $(G(t))$ defined in (11). Therefore, for $\varepsilon = 1/\sqrt{N}$, $\sqrt{N}(Z_N(t) - x(t))_{t\geq 0}$ and $\varepsilon^{-1}(X_\varepsilon(t) - x(t))_{t\geq 0}$ converge to a Gaussian process having the same distribution.

For sake of clarity, we have detailed above these approximations for diffusion processes with drift and diffusion coefficients depending only on the state space variable x. However, this expansion of $X_\varepsilon(.)$ is still true for time-dependent jump processes leading to diffusion processes $(X_N(.))$ with drift term and diffusion matrix respectively $b(t,x)$ and $\frac{1}{N}\Sigma(t,x)$: the corresponding ODE and Gaussian process are respectively,

$$x(t) = x_0 + \int_0^t b(u,x(u))du, \quad g(t) = \int_0^t \Phi(t,s)\sigma(s,x(s))dB(s). \tag{16}$$

2.2. *The diffusion approximation applied to the SIR epidemic model*

We apply first the generic method leading successively to $b(.)$, $\Sigma(.)$ and (X_N) described in 2.1 to the *SIR* model introduced in section 1 and in Britton and Giardina (2014) through the two-dimensional continuous-time Markov jump process $Z(t) = (S(t), I(t))$, to build the associated *SIR* diffusion process. Along to its initial state $Z(0) = (S(0), I(0))$, the Markov jump process is characterized by two transitions, $(S,I) \xrightarrow{\frac{\lambda}{N}SI} (S-1, I+1)$ and $(S,I) \xrightarrow{\gamma I} (S, I-1)$. Parameters λ and $\gamma = 1/d$ represent the transmission rate and the recovery rate (or the inverse of the mean infection duration d), respectively. The rate $\lambda SI/N$ translates two main assumptions: the population mix homogeneously (same λ for each pair of one S and one I) and the transmission is proportional to the fraction of infectious individuals in the population, I/N (frequency-dependent formulation of the transmission term).

The diffusion approximation of the initial jump process describing the epidemic dynamics can be summarized by three steps.
First, transition rates need to be defined. The original *SIR* jump process in a closed population has state space $\{0,\ldots,N\}^2$, the possible jumps (symbolized by l in 2.1) are $(-1,1)$ and $(0,-1)$ with transition rates,

$$q_{(S,I),(S-1,I+1)} = \lambda S\,\frac{I}{N} = \alpha_{(-1,1)}(S,I); \quad q_{(S,I),(S,I-1)} = \gamma I = \alpha_{(0,-1)}(S,I).$$

Second, normalizing $Z(t)$ by the population size N, we obtain, setting $y = (s,i) \in [0,1]^2$,

$$\frac{1}{N}\alpha_{(-1,1)}([Ny]) \underset{N\to+\infty}{\longrightarrow} \beta_{(-1,1)}(s,i) = \lambda si; \quad \frac{1}{N}\alpha_{(0,-1)}([Ny]) \underset{N\to+\infty}{\longrightarrow} \beta_{(0,-1)}(s,i) = \gamma i.$$

These two limiting functions clearly satisfy (H1)-(H2).
Finally, the two functions given in (6), (7) are well defined and now depend on (λ,γ).
Set $\theta = (\lambda,\gamma)$ and denote $b(\theta,y)$ and $\Sigma(\theta,y)$ the associated functions. We get

$$b(\theta,(s,i)) = \begin{pmatrix} -\lambda si \\ \lambda si - \gamma i \end{pmatrix}; \quad \Sigma(\theta,(s,i)) = \begin{pmatrix} \lambda si & -\lambda si \\ -\lambda si & \lambda si + \gamma i \end{pmatrix}. \tag{17}$$

 ISSN: 2102-6238

We can now derive the diffusion approximation of the *SIR* epidemics obtained in (13). On a filtered probability space $(\Omega, \mathscr{F} = (\mathscr{F}_t), \mathbb{P})$, let $B(t) = {}^t(B_1(t), B_2(t))$ denote a standard two-dimensional Brownian motion and consider the Choleski decomposition of $\Sigma(\theta, y)$,

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

Assume that $Z(0) = (S(0), I(0))$ satisfies $(\frac{S(0)}{N}, \frac{I(0)}{N}) \underset{N\to+\infty}{\longrightarrow} (s_0, i_0) = x_0$. Then, $X_N(t) = \begin{pmatrix} S_N(t) \\ I_N(t) \end{pmatrix}$ satisfies the stochastic differential equation defined by $X_N(0) = x_0$ and

$$\begin{aligned} dS_N(t) &= -\lambda S_N(t) I_N(t) dt + \tfrac{1}{\sqrt{N}} \sqrt{\lambda S_N(t) I_N(t)} dB_1(t), \\ dI_N(t) &= (\lambda S_N(t) I_N(t) - \gamma I_N(t)) dt - \tfrac{1}{\sqrt{N}} \left(\sqrt{\lambda S_N(t) I_N(t)} dB_1(t) - \sqrt{\gamma I_N(t)} dB_2(t) \right). \end{aligned}$$

In order to visualize the influence of the population size N on the sample paths of the normalized jump process $Z_N(t) = Z(t)/N$, several trajectories have been simulated using an *SIR* model with parameters $(\lambda, \gamma) = (0.5, 1/3)$. Results are displayed in Figure 1. We observe that, as the population size increases, the stochasticity of sample paths decreases. However, it still keeps a non negligible stochasticity for a large population size ($N = 10000$). Since the peak of $I_N(t)$ is quite small (about 0.08 here), this can be explained by a moderate size of the ratio "signal over noise" even for large N (here of order $0.08/0.01$).

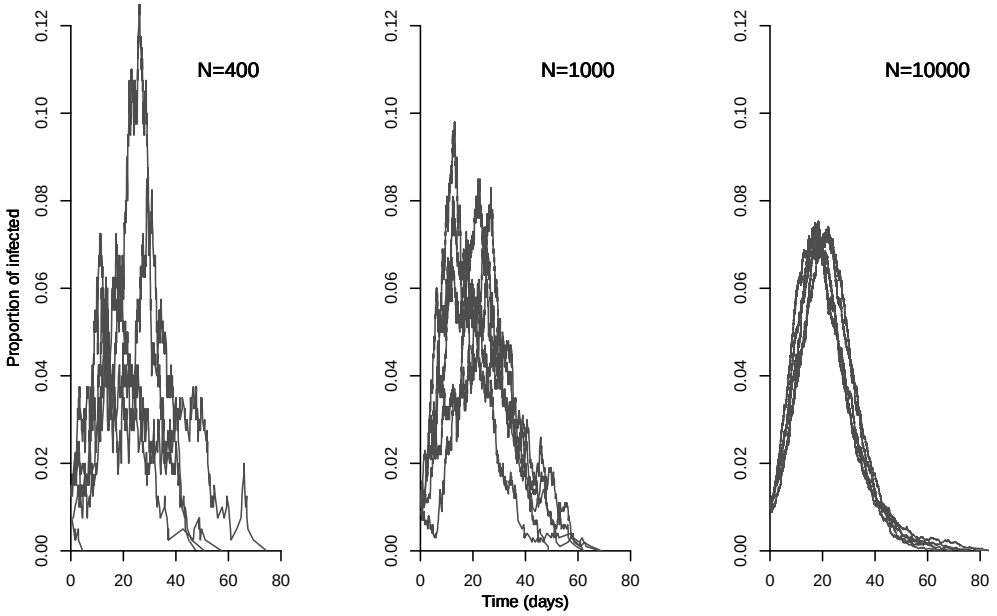

FIGURE 1. *Five simulated trajectories of the proportion of infected individuals over time using the SIR Markov jump process for* $(s(0), i(0)) = (0.99, 0.01)$ $(\lambda, \gamma) = (0.5, 1/3)$ *and for each* $N = \{400, 1000, 10000\}$ *(from left to right).*

 ISSN: 2102-6238

2.3. *The diffusion approximation applied to the SIRS epidemic model with seasonal forcing*

Another important class of epidemics models introduced in section 1 is the *SIRS* model, which allows possible reinsertion of removed individuals into S class. The additional transition writes as $(S,I) \overset{\delta(N-S-I)}{\longrightarrow} (S+1,I)$, where δ is the average rate of immunity waning. To mimic recurrent epidemics, additional mechanisms need to be considered. Hence, an appropriate model to describe recurrent epidemics is the *SIRS* model with seasonal transmission (at rate $\lambda(t)$), external immigration flow in the I class (at rate η) and, when the time-scale of study is large, demography (with birth and death rates equal to μ for a stable population of size N). Seasonality in transmission is captured using a time non homogeneous transmission rate, expressed under a periodic form

$$\lambda(t) := \lambda_0(1+\lambda_1 sin(2\pi t/T_{per})) \tag{18}$$

where λ_0 is the baseline transition rate, λ_1 the amplitude of the seasonality in transmission and T_{per} the period of the seasonal trend (see Keeling and Rohani, 2011, Chapter 5).

Figure 2 illustrates the dynamics of the *SIRS* model (in ODE formalism) which is forced using sinusoidal terms. In particular, given the parameter values we have chosen, we can notice two distinct regimes: one with annual cycles (top panel) and the other with biennial dynamics (middle panel). The qualitative changes in model dynamics following the modification of a control parameter or *bifurcation parameter* (here λ_1) are summarized in the *bifurcation diagram* (bottom panel; more detail in the figure caption).

The diffusion approximation is built following the same generic scheme of section 2.1 as for the *SIR* model in section 2.2. Assuming again a constant population size, we obtain a new two dimensional system with four transitions for the corresponding Markov jump process. The jump process is time-dependent and so we have to check (H1b)-(H2b). First, four jumps l corresponding to functions α_l are possible in this model, $l \in \{(-1,1);(-1,0);(0,-1);(1,0)\}$. The corresponding transition rates are $q_{(S,I),(S-1,I+1)}$, $q_{(S,I),(S-1,I)}$, $q_{(S,I),(S,I-1)}$ and $q_{(S,I),(S+1,I)}$, which respectively write as

$(S,I) \overset{\frac{\lambda(t)}{N}S(I+N\eta)}{\longrightarrow} (S-1,I+1) \Rightarrow \alpha_{(-1,1)} = \frac{\lambda(t)}{N}S(I+N\eta)$

$(S,I) \overset{\mu S}{\longrightarrow} (S-1,I) \quad \Rightarrow \alpha_{(-1,0)} = \mu S$

$(S,I) \overset{(\gamma+\mu)I}{\longrightarrow} (S,I-1) \Rightarrow \alpha_{(0,-1)} = (\gamma+\mu)S$

$(S,I) \overset{\mu N+\delta(N-S-I)}{\longrightarrow} (S+1,I) \Rightarrow \alpha_{(1,0)} = \mu N + \delta(N-S-I)S.$

Second, straightforward computations yield, for $(s,i) \in [0,1]^2$,

$\beta_{(-1,1)}(t,(s,i)) = \lambda(t)s(i+\eta)$

$\beta_{(-1,0)}(t,(s,i)) = \mu s$

$\beta_{(0,-1)}(t,(s,i)) = (\gamma+\mu)i$

$\beta_{(1,0)}(t,(s,i)) = \mu + \delta(1-s-i).$

Clearly, (H1b) and (H2b) are satisfied.

Finally, setting $\theta = (\lambda_0,\lambda_1,\gamma,\delta,\eta,\mu)$, the associated drift function $b(\theta,(s,i))$ and diffusion

 ISSN: 2102-6238

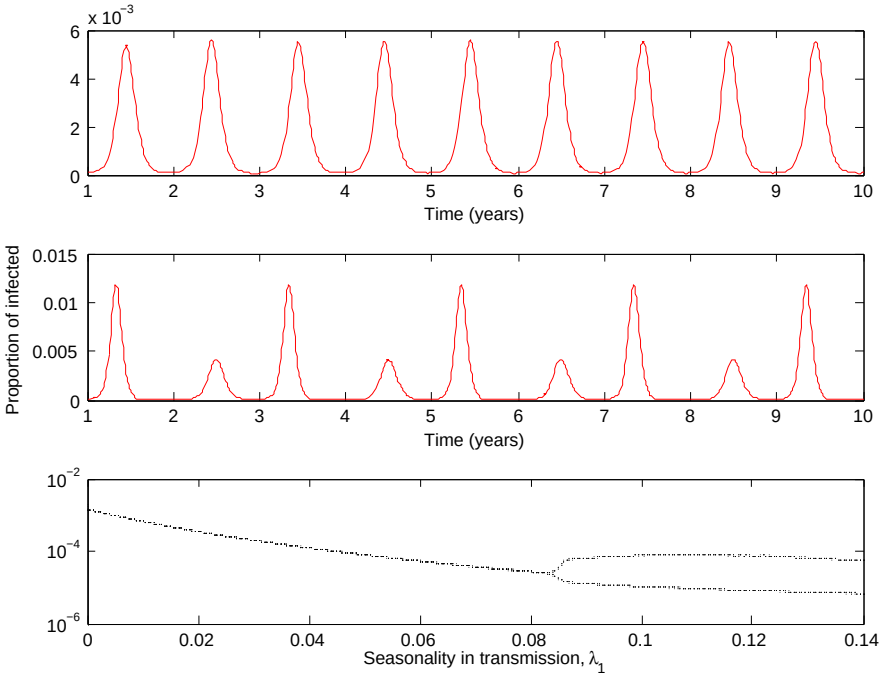

FIGURE 2. *Proportion of infected individuals, $I(t)$, over time simulated using the ODE variant of the SIRS model with $N = 10^7$, $T_{per} = 365$, $\mu = 1/(50 \times T_{per})$, $\eta = 10^{-6}$, $(s_0, i_0) = (0.7, 10^{-4})$ and $(\lambda_0, \gamma, \delta) = (0.5, 1/3, 1/(2 \times 365))$. The top panel corresponds to $\lambda_1 = 0.05$, the middle panel to $\lambda_1 = 0.1$. The bottom panel represents the bifurcation diagram with respect to λ_1. This diagram is constructed as follows: for each specific value of λ_1, the proportion of infected individuals (once the dynamics reached the asymptotic state) is plotted at the last day of each year. For λ_1 values smaller than 0.08, the simulated dynamics is annual (as the forcing function in (18)) and one dot on y-axis corresponds to each value of λ_1. For values of λ_1 above 0.08, biennial dynamics occur, which is shown by the two dots on the y-axis for each value of λ_1.*

matrix $\Sigma(\theta, t, (s,i))$ are

$$b(\theta, t, (s,i)) = \begin{pmatrix} -\lambda(t)s(i+\eta) + \delta(1-s-i) + \mu(1-s) \\ \lambda(t)s(i+\eta) - (\gamma+\mu)i \end{pmatrix}, \tag{19}$$

$$\Sigma(\theta, t, (s,i)) = \begin{pmatrix} \lambda(t)s(i+\eta) + \delta(1-s-i) + \mu(1+s) & -\lambda(t)s(i+\eta) \\ -\lambda(t)s(i+\eta) & \lambda(t)s(i+\eta) + (\gamma+\mu)i \end{pmatrix}. \tag{20}$$

Choosing $\sigma(\theta, t, (s,i))$ such that $\sigma(.)^t\sigma(.) = \Sigma(\theta, t, (s,i))$, we obtain that the approximating diffusion $X_N(t)$ satisfies

$$dX_N(t) = b(\theta, t, (S_N, I_N))dt + \frac{1}{\sqrt{N}}\sigma(\theta, t, (S_N, I_N)); \; X_N(0) = x_0. \tag{21}$$

3. Inference for diffusion processes with small diffusion coefficient: general framework

Inference for diffusion processes observed on a finite time-interval presents some specific properties. For sake of clarity, a short recap of classical results for diffusion processes inference is then

 ISSN: 2102-6238

given. We first present the general framework required for time-dependent diffusion processes and then detail these results.

3.1. *Framework and main assumptions*

Our concern here is parametric inference. In the sequel, we assume that the parameter set Θ is a subset of $\mathbb{R}^a \times \mathbb{R}^b$ and denote $\theta = (\alpha, \beta) \in \Theta$ the current parameter.
In order to deal with general epidemics, we consider time-dependent diffusion processes on $\mathbb{R}^p$ with small diffusion coefficient $\varepsilon = 1/\sqrt{N}$ satisfying the following SDE:

$$dX(t) = b(\alpha;t,X(t))dt + \varepsilon\sigma(\beta;t,X(t))\, dB(t)\,; \quad X(0) = x_0, \tag{22}$$

where $(B(t)_{t\geq 0})$ is a p-dimensional standard Brownian motion defined on a probability space $\mathbb{P} = (\Omega, (\mathscr{F}_t)_{t\geq 0}, P)$, $b(\alpha;t,.) : \mathbb{R}^p \to \mathbb{R}^p$ and $\sigma(\beta;t,.) : \mathbb{R}^p \to \mathbb{R}^p \times \mathbb{R}^p$.
We assume that $b(\alpha;t,x)$ and $\sigma(\beta;t,x)$ are measurable in (t,x), Lipschitz continuous with respect to the second variable and satisfy a linear growth condition for all $t \geq 0, x, y \in \mathbb{R}^p$, *i.e.* there exists a global constant K such that

(S1): $\forall(\alpha,\beta) \in \Theta, \| b(\alpha;t,x) - b(\alpha;t,y) \| + \| \sigma(\beta;t,x) - \sigma(\beta;t,y) \| \leq K \| x - y \|$;
(S2): $\forall(\alpha,\beta) \in \Theta, \| b(\alpha;t,x) \|^2 + \| \sigma(\beta;t,x) \|^2 \leq K(1 + \| x \|^2)$;
(S3): $\forall(\beta,t,x)$, $\Sigma(\beta;t,x) = \sigma(\beta;t,x)^t\sigma(\beta;t,x)$ is invertible.

Assumptions (S1)-(S3) are classical assumptions that ensure that for all θ equation (22) admits a unique strong solution (see e.g. Karatzas and Shreve, 1998, Chapter 5.2.B.). Under some additional regularity assumptions, the expansion (14) still holds for time-dependent diffusion processes (Azencott, 1982).
Introducing the dependence with respect to t and parameters (α,β) in (8), (9), (15) yields,

$$\begin{array}{rl} x_\alpha(t) = & b(\alpha;s,x_\alpha(s))ds, \\ g_{\alpha,\beta}(t) = & \int_0^t \Phi_\alpha(t,s)\sigma(\beta;s,x_\alpha(s))dB(s), \quad \text{with } \Phi_\alpha \text{ such that} \\ \frac{\partial\Phi_\alpha}{\partial t}(t,u) = & \frac{\partial b}{\partial x}(\alpha;t,x_\alpha(t))\Phi_\alpha(t,u)\ ,\ \Phi_\alpha(u,u) = I_p, \end{array} \tag{23}$$

where $\frac{\partial b}{\partial x}(\alpha;t,x)$is the $p \times p$ matrix $(\frac{\partial b_i}{\partial x_j}(\alpha;t,x))_{ij}$. Then the expansion (14) writes

$$\begin{array}{rl} X_N(t) = & X_\varepsilon(t) = x_\alpha(t) + \varepsilon g_{\alpha,\beta}(t) + \varepsilon^2 R_{\varepsilon,\alpha,\beta}(t), \quad \text{where} \\ sup_{t\leq T} & \| \varepsilon R_{\varepsilon,\alpha,\beta}(t) \| \to 0 \text{ in probability as } \varepsilon \to 0. \end{array} \tag{24}$$

Let θ_0 be the true value of the parameter. For inference, we need the following assumptions
(S4): $\Theta = K_a \times K_b$ is a compact set of $\mathbb{R}^{a+b}$, $\theta_0 \in \mathring{\Theta}$;
(S5): For all $t \geq 0$, $b(\alpha;t,x) \in C^2(K_a \times \mathbb{R}^p, \mathbb{R}^p)$ and $\sigma(\beta;t,x) \in C^2(K_a \times \mathbb{R}^p, M_p(\mathbb{R}))$;
(S6): $\alpha \neq \alpha' \Rightarrow b(t;\alpha,x_\alpha(t)) \not\equiv b(t;\alpha',x_{\alpha'}(t))$;
(S7): $\beta \neq \beta' \Rightarrow \Sigma(t;\beta,x_{\alpha_0}(t)) \not\equiv \Sigma(t;\beta',x_{\alpha_0}(t))$.

Finally, we denote by $\mathbb{P}_\theta^\varepsilon = \mathbb{P}_{\alpha,\beta}^\varepsilon$ the distribution of $(X(t))$ satisfying (22) on $C = (C([0,T],\mathbb{R}^p), \mathscr{C})$, with $\mathscr{C}$ the Borel σ-algebra on $C([0,T],\mathbb{R}^p)$.

 ISSN: 2102-6238

3.2. *Short recap of inference for diffusion processes*

Classical results concern autonomous diffusion processes, i.e. with drift and diffusion coefficients $b(\alpha,x)$ and $\sigma(\beta,x)$. They differ according to the observations.

(1) Continuous observation of the sample path on $[0,T]$
Consider first the case where the diffusion $(X(t))$ with drift term $b(.)$ and diffusion matrix $\sigma(.)$ is continuously observed on a finite time interval $[0,T]$. Let $\mathbb{P}^T_{b,\sigma}$ denote its distribution on $(C,\mathscr{C})$. Then, the two distributions $\mathbb{P}^T_{b,\sigma}$ and $\mathbb{P}^T_{b,\sigma'}$ are orthogonal if $\sigma(.) \not\equiv \sigma'(.)$. (see e.g Lipster and Shiryaev, 2001). Therefore, this excludes maximum likelihood approaches for different values of β. The consequence is that it is possible to identify the diffusion coefficient (or parameters present in it) without a statistical method. This is done for instance using that the quadratic variations of $X(t)$, $\sum_{k=1}^n (X(t_k)-X(t_{k-1}))^2$, tend to $\int_0^T \sigma^2(X(s))ds$ in probability, as Δ tends to 0 (where $t_k = k\Delta$ and $T = n\Delta$).
Assuming that $\beta = \beta_0$ is fixed, two diffusion processes with drift functions $b(\alpha,x)$, $b(\alpha',x)$ have absolutely continuous distributions on $(C([0,T],\mathbb{R}^p),\mathscr{C})$. Then, (see Kutoyants, 1984), the maximum likelihood estimator $\hat{\alpha}_\varepsilon$ is consistent and satisfies that, under $\mathbb{P}^{\varepsilon,T}_{\alpha_0,\beta_0}$, as $\varepsilon \to 0$,

$$\varepsilon^{-1}(\hat{\alpha}_\varepsilon - \alpha_0) \to \mathscr{N}\left(0, I_b^{-1}(\alpha_0,\beta_0)\right) \text{ in distribution, with} \tag{25}$$

$$I_b(\alpha_0,\beta_0) = \left(\int_0^T {}^t \frac{\partial b}{\partial \alpha_i}(\alpha_0, x_{\alpha_0})(s) \Sigma^{-1}(\beta_0, x_{\alpha_0}(s)) \frac{\partial b}{\partial \alpha_j}(\alpha_0, x_{\alpha_0})(s) \right)_{1\le i,j\le a} \tag{26}$$

The matrix $I_b(\alpha_0,\beta_0)$ is the Fisher information of this statistical experiment, which has to be assumed invertible for getting (25).

(2) Discrete observations of the sample path with sampling interval Δ **on** $[0,T]$
Assume that $T = n\Delta$ and that $X(0) = x_0$ is fixed. Then, the observations consist of the n-tuple $(X(k\Delta), k = 1,\dots,n)$. Denote by $\mathbb{P}^{\varepsilon,\Delta}_{\alpha,\beta}$ the distribution on $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ of the n-tuple $(X(k\Delta), k = 1,\dots,n)$. Note that now, distributions corresponding to different values of (α,β) are absolutely continuous. The main difficulty here lies in the intractable likelihood. This is a well known problem for discrete observations of diffusion processes. Alternative approaches based on M-estimators or contrast processes (see e.g. van der Vaart, 2000) have to be investigated. There are distinct asymptotic results according to Δ.
High frequency sampling: $\Delta = \Delta_n \to 0$ with $T = n\Delta_n$
This implies that the number of observations $n \to \infty$. Then, Gloter and Sørensen (2009) proved, under assumptions linking the two asymptotics ε and n, the existence of consistent and asymptotically Gaussian estimators $(\tilde{\alpha}_{\varepsilon,n}, \tilde{\beta}_{\varepsilon,n})$ of (α,β), which converge at different rates, parameters in the drift function being estimated at rate ε^{-1} and parameters in the diffusion coefficient at rate $\sqrt{n} = \Delta_n^{-1/2}$:

$$\begin{pmatrix} \varepsilon^{-1}(\hat{\alpha}_{\varepsilon,n} - \alpha_0) \\ \sqrt{n}(\hat{\beta}_{\varepsilon,n} - \beta_0) \end{pmatrix} \underset{n\to\infty,\varepsilon\to 0}{\longrightarrow} \mathscr{N}\left(0, \begin{pmatrix} I_b^{-1}(\alpha_0,\beta_0) & 0 \\ 0 & I_\sigma^{-1}(\alpha_0,\beta_0) \end{pmatrix}\right). \tag{27}$$

 ISSN: 2102-6238

The matrix I_b is the matrix defined in (26), the matrix I_σ is

$$I_\sigma(\alpha,\beta) = \left(\frac{1}{2T}\int_0^T Tr(\frac{\partial\Sigma}{\partial\beta_k}\Sigma^{-1}\frac{\partial\Sigma}{\partial\beta_l})(\beta;s,x_\alpha(s))ds\right)_{1\le k,l\le b}, \tag{28}$$

where $I_b(\alpha_0,\beta_0)$ and $I_\sigma(\alpha_0,\beta_0)$ are assumed invertible ($Tr(A)$ denoting the trace of a matrix A).
Low frequency sampling: fixed Δ
Since $T = n\Delta$ is finite, this implies that the number of observations n is finite. Investigating the inference for a finite number of observations, n, is not classical. But this occurs in practice for epidemics, along with the population size asymptotics (*i.e.* $\varepsilon \to 0$). This is a framework that we investigated in Guy et al. (2014) and that will be summarised in section 4.2.

4. Inference for discretely observed diffusions motivated by epidemic data

This section sums up the statistical results obtained in Guy et al. (2014) and Guy et al. (2015) for multidimensional time-dependent diffusion processes $(X(t), 0 \le t \le T)$ with small diffusion matrix, obtained as approximations of epidemic processes. Observations consist of discrete observations of the sample path with sampling Δ on a fixed time interval $T = n\Delta$. In the sequel, we denote $(X(t_k))$ the n-tuple $(X(t_1),\dots,X(t_n))$ with $t_i = i\Delta$.
At first glance, the low frequency sampling seems a priori more appropriate for epidemic data. However, both high and low frequency observations could be appropriate in practice, because the choice of the statistical framework depends more on the relative magnitudes between T, Δ and the population size N ($= \varepsilon^{-1/2}$) than on their accurate values.
We present successively results obtained for the high frequency sampling, where the asymptotics is $\varepsilon = 1/\sqrt{N} \to 0$, $\Delta_n = T/n \to 0$ (section 4.1), and for the low frequency sampling, $\varepsilon = 1/\sqrt{N} \to 0$, Δ (and also the number of observations, n), fixed (section 4.2).

4.1. *High frequency observations*

We assume that both ε and Δ go to 0. The sample path $(X(t))$ is observed at times $t_k = k\Delta$. Hence, the number of observations n goes to infinity. The inference in the case $\sigma(\beta,x) \equiv 1$ was first investigated for one dimensional diffusion process by Genon-Catalot (1990), using expansion (14).
The results obtained in Gloter and Sørensen (2009) require conditions linking ε and Δ that do not fit epidemic data, where generally the parameter ε is small, and orders of magnitude for N and n satisfy $N >> n$ so that Δ is comparatively large with respect to ε.
We proposed in Guy et al. (2014) another method based on the Taylor expansion of $X_\varepsilon(.)$, which extends results obtained in Genon-Catalot (1990). This yields another contrast process based on the property that the random variables

$$B_k(\alpha,X) = X(t_k) - x_\alpha(t_k) - \Phi_\alpha(t_k,t_{k-1})\left[X(t_{k-1}) - x_\alpha(t_{k-1})\right] \tag{29}$$

are approximately conditionally independent centered Gaussian random variables on $\mathbb{R}^p$ with covariance matrix

$$S_k(\alpha,\beta) = \frac{1}{\Delta}\int_{t_{k-1}}^{t_k} \Phi_\alpha(t_k,s)\Sigma(\beta;s,x_\alpha(s))\,{}^t\Phi_\alpha(t_k,s)ds. \tag{30}$$

 ISSN: 2102-6238

Let

$$U_{\varepsilon,\Delta}(\alpha,\beta;(X(t_k))) = \sum_{k=1}^{n}\left(log(det\ S_k(\alpha,\beta)) + \frac{1}{\varepsilon^2\Delta}{}^tB_k(\alpha,X)\ S_k^{-1}(\alpha,\beta)\ B_k(\alpha,X)\right). \quad (31)$$

Under $\mathbb{P}^{\varepsilon}_{\alpha_0,\beta_0}$, as $\varepsilon,\ \Delta\to 0$, we proved in Guy et al. (2014) that $\varepsilon^2 U_{\varepsilon,\Delta}(\alpha,\beta;(X(t_k))) \to K(\alpha_0,\alpha;\beta)$, with $K(\alpha_0,\alpha;\beta) = \int_0^T {}^t\Gamma(\alpha_0,\alpha;t)\Sigma^{-1}(\beta,x_{\alpha_0}(t)\Gamma(\alpha_0,\alpha;t)dt$ and
$\Gamma(\alpha_0,\alpha;t) = b(\alpha_0;t,x_{\alpha_0}(t)) - b(\alpha;t,x_\alpha(t)) - \frac{\partial b}{\partial x}(\alpha;t,x_\alpha(t))(x_{\alpha_0}(t)-x_\alpha(t))$.
Define the minimum contrast estimator

$$(\hat\alpha_{\varepsilon,\Delta},\hat\beta_{\varepsilon,\Delta}) = \underset{\alpha,\beta}{argmin}\ U_{\varepsilon,\Delta}(\alpha,\beta;(X(t_k)))\,.$$

We proved for autonomous (Guy et al., 2014) and time-dependent (Guy et al., 2015) diffusion processes the following result (for proofs, see these two articles cited).

Theorem 4.1. *Assume (S1)-(S7). Then, under $\mathbb{P}^{\varepsilon}_{\alpha_0,\beta_0}$, as $\varepsilon,\ n\to\infty$ ($\Delta = T/n\to 0$),*

$$\begin{pmatrix}\varepsilon^{-1}(\hat\alpha_{\varepsilon,\Delta}-\alpha_0)\\ \sqrt{n}(\hat\beta_{\varepsilon,\Delta}-\beta_0)\end{pmatrix} \underset{n\to\infty,\varepsilon\to 0}{\longrightarrow} \mathcal{N}\left(0,\begin{pmatrix} I_b^{-1}(\alpha_0,\beta_0) & 0\\ 0 & I_\sigma^{-1}(\alpha_0,\beta_0)\end{pmatrix}\right) \quad \textit{in distribution,} \quad (32)$$

where the two matrices $I_b(\alpha,\beta)$ and $I_\sigma(\alpha,\beta)$ are defined in (26), (28).

Remark 1. *The covariance matrix $S_k(\alpha,\beta)$ and the random variables $B_k(\alpha,X)$ depend on $\Phi_\alpha(t,s)$ and might be difficult to compute. Using that, for small Δ and ε, $S_k(\alpha,\beta)\simeq\Sigma(\beta;t_{k-1},X(t_{k-1}))$ and $\Phi_\alpha(t_k,t_{k-1})\simeq I_p+\Delta\frac{\partial b}{\partial x}(\alpha,t_{k-1},x(t_{k-1}))$, we can replace $S_k(\alpha,\beta)$ and $\Phi_\alpha(t_k,t_{k-1})$ by these latter more tractable quantities in (31). This yields another contrast process $\tilde U_{\varepsilon,\Delta}(\alpha,\beta;(X(t_k)))$, together with another minimum contrast estimators $(\tilde\alpha_{\varepsilon,\Delta},\tilde\beta_{\varepsilon,\Delta})$, which are asymptotically equivalent to $(\hat\alpha_{\varepsilon,\Delta},\hat\beta_{\varepsilon,\Delta})$.*

4.2. Low frequency observations

The sampling interval Δ is fixed, so that the data consist of a finite number n of observations $(X(t_k),k=1,\dots,n)$ with $T=n\Delta$. Only parameters in the drift function can be consistently estimated. This agrees with the previous results where the rate of estimation for β in the high frequency set-up is $\sqrt{n}$. This can be easily illustrated considering the Brownian motion with drift,

$$dX(t) = \alpha dt + \varepsilon\ \beta\ dB(t); X(0) = 0.$$

The random variables $(X(t_k)-X(t_{k-1}),k=1,\dots,n)$ form a sequence of Gaussian independent random variables, $\mathcal{N}(\alpha\Delta,\varepsilon^2\sigma^2\Delta)$. The likelihood is explicit and the maximum likelihood estimators are $\hat\alpha_{\varepsilon,\Delta} = \frac{X(T)}{T}$, $\quad \hat\beta^2_{\varepsilon,\Delta} = \frac{1}{n\Delta\varepsilon^2}\sum_1^n(X(t_k)-X(t_{k-1})-\Delta\hat\alpha_{\varepsilon,\Delta})^2$.
Under $\mathbb{P}^{\varepsilon}_{\theta_0}$, $\varepsilon^{-1}(\hat\alpha_{\varepsilon,\Delta}-\alpha_0) = \beta_0\frac{B(T)}{T}$, which has distribution $\mathcal{N}(0,\beta_0^2/T)$, and therefore $\hat\alpha_{\varepsilon,\Delta}$ is consistent and Gaussian as $\varepsilon\to 0$. The estimator of $\hat\beta^2_{\varepsilon,\Delta}$ satisfies $\hat\beta^2_{\varepsilon,\Delta} = \beta_0^2(\frac{1}{n}\sum_1^n U_k^2 - \frac{1}{n}\frac{B(T)^2}{T})$, where (U_k) are i.i.d $\mathcal{N}(0,1)$. Hence, since n is fixed, $\hat\beta^2_{\varepsilon,\Delta}$ is a given random variable which does

 ISSN: 2102-6238

not depend on ε. Its expectation is $\beta_0^2(1-\frac{1}{n}) \neq \beta_0^2$, implying that it is a biased estimator of β_0.

Therefore, for estimating α in presence of unknown parameters in the diffusion coefficient, a conditional least squares method based on the random variables $B_k(\alpha)$ defined in (29) must be used. It results in substituting $S_k(\alpha,\beta)$ by I_p in (31). Consistent and ε^{-1} convergent estimators of α are obtained (see Guy et al., 2014).

However, for epidemic data, we can use the property that the same parameters both appear in the drift and diffusion coefficients of the approximating diffusion processes. Therefore, substituting β by α in (31) yields another contrast process which writes, using (29), (30),

$$U_{\varepsilon,\Delta}(\alpha;(X(t_k))) = \sum_{k=1}^{n} \left(log(det\ S_k(\alpha,\alpha)) + \frac{1}{\varepsilon^2\Delta}\, {}^tB_k(\alpha,X)\, S_k^{-1}(\alpha,\alpha)\, B_k(\alpha,X) \right). \tag{33}$$

Then, under $\mathbb{P}^{\varepsilon}_{\theta_0}$, as $\varepsilon \to 0$,

$$\varepsilon^2 U_{\varepsilon,\Delta}(\alpha,(X(t_k))) \to K_\Delta(\alpha_0,\alpha) = \frac{1}{\Delta}\sum_{1}^{n} {}^tB_k(\alpha,x(\alpha_0,.))S_k^{-1}(\alpha,\alpha)B_k(\alpha,x(\alpha_0,.))\ a.s.$$

The identifiability assumption ensuring that the function $\alpha \to K_\Delta(\alpha_0,\alpha)$ is positive and equal 0 for $\alpha = \alpha_0$ leads to the following modification of (S6)

(S6b): $\alpha \neq \alpha' \Rightarrow (x_\alpha(t_k), k = 1\ldots,n) \not\equiv (x_{\alpha'}(t_k), k = 1\ldots,n)$.

The minimum contrast estimator $\hat{\alpha}_{\varepsilon,\Delta}$ is consistent and satisfies that, under $\mathbb{P}_{\alpha_0,\varepsilon}$,

$$\varepsilon^{-1}(\hat{\alpha}_{\varepsilon,\Delta} - \alpha_0) \underset{\varepsilon\to 0}{\longrightarrow} \mathcal{N}(0, I_\Delta^{-1}(\alpha_0)). \tag{34}$$

The matrix $I_\Delta(\alpha_0)$ is explicit and satisfies that, as $\Delta \to 0$, $I_\Delta(\alpha_0) \to I_b(\alpha_0)$, the Fisher information of the continuous observations model of $(X(t), t \leq T)$ defined in (26). As expected, this property is not recovered when using conditional least squares (used in the general case of β unknown) which leads to a covariance matrix $V_\Delta(\theta_0)$ for $\tilde{\alpha}_{\varepsilon,\Delta}$, tending to $(J(\theta_0)^{-1}I_b(\theta_0)J(\theta_0)^{-1})$ instead of $I_b^{-1}(\theta_0)$ as $\Delta \to 0$. The matrix $J(\theta)$ is explicit and given in Proposition 3.1 of Guy et al. (2014).

5. Statistical inference for partially observed epidemic dynamics

In the case of epidemics, numbers of susceptible and infected individuals over time are generally not observed. In practice, (sometimes noisy) observations are often assumed to correspond to aggregated numbers, over the sampling interval Δ, of newly infected individuals (i.e. $\int_{t_{k-1}}^{t_k} \lambda S(s)I(s)ds$). In the *SIR* diffusion model, this corresponds to the recovered individuals $(R(t_k) - R(t_{k-1})), k = 1,\ldots n$ (for diseases with short duration of the infected period). Hence, this situation can be assimilated, as a first attempt, to the case where only one coordinate can be observed.

 ISSN: 2102-6238

In this section, we consider the case of a two-dimensional diffusion process $X(t) = {}^t(Y(t), Z(t))$ where only the first coordinate $Y(t)$ is discretely observed on a fixed time interval $[0,T]$ with sampling Δ. Therefore, the observations are now

$$Y(t_k),\ k = 1, \dots n, \quad \text{with } t_k = k\Delta, \quad T = n\Delta. \tag{35}$$

For continuous observations of $(Y(t))$ on a finite time interval $[0,T]$, two studies (James and Le Gland, 1995; Kutoyants, 1994) are concerned with parametric inference in this statistical framework. Both studied the maximum likelihood estimator of parameters in the drift function for a diffusion matrix equal to $\varepsilon^2 I_p$. This likelihood is difficult to compute since it relies on integration on the unobserved coordinate. James and Le Gland (1995); Kutoyants (1994) proposed filtering approaches to compute this likelihood, as it is done for general Hidden Markov Models (see e.g. Cappé et al., 2005; Douc et al., 2011). Here, we can take advantage of the presence of ε and extend to partial observations the method by contrast processes and M- estimators that had been developed for complete observations (Genon-Catalot, 1990; Sørensen and Uchida, 2003; Gloter and Sørensen, 2009; Guy et al., 2014).

We study the case of small (or high frequency) sampling interval, $\Delta = \Delta_n \to 0$, on a fixed time interval $[0,T]$ with $T = n\Delta$, which yields explicit results. This allows to disentangle problems coming from discrete observations and those coming from the missing observation of one coordinate and hence provides a better understanding of the problems rising in this context. The case of Δ fixed could be studied similarly, with more cumbersome notations and no such insights.
First, the notations required are introduced, results are then stated, and finally, to illustrate this approach, the example of a two-dimensional Ornstein-Uhlenbeck process, where all the computations are explicit is developed.

5.1. Notations for partial observations and assumptions

Some specific notations need to be introduced. For $x \in \mathbb{R}^2$, $X_\varepsilon(t)$, the diffusion process, $B(t)$ the Brownian motion, and M a 2×2 matrix, we write

$$x = \begin{pmatrix} y \\ z \end{pmatrix};\ X_\varepsilon(t) = \begin{pmatrix} Y_\varepsilon(t) \\ Z_\varepsilon(t) \end{pmatrix};\ B(t) = \begin{pmatrix} B_1(t) \\ B_2(t) \end{pmatrix};\ M = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}. \tag{36}$$

For functions defined for $x \in \mathbb{R}^2$ and depending on parameters, we use two distinct notations.
Derivation with respect to the state variable x is $\frac{\partial f}{\partial x} = \begin{pmatrix} \frac{\partial f_1}{\partial y} & \frac{\partial f_1}{\partial z} \\ \frac{\partial f_2}{\partial y} & \frac{\partial f_2}{\partial z} \end{pmatrix}$.
For $\theta = (\theta_i)$, derivation with respect θ_i is $\nabla_i f$ with the convention that, for $f(\theta, x(\theta,t))$,
$\nabla_i f(\theta, x(\theta,t)) = \frac{\partial f}{\partial \theta_i}(\theta, x(\theta,t)) + \frac{\partial f}{\partial x}(\theta, x(\theta,t)) \frac{\partial x}{\partial \theta_i}(\theta,t)$.
We keep the notation $\frac{\partial f}{\partial \theta_i}$ for the derivation w.r.t. θ_i if needed.
We also use the notation $\nabla_{ij} f = \nabla_i(\nabla_j f)$.
As in the previous section, the starting point of the diffusion is assumed to be fixed, $X(0) = x_0 = \begin{pmatrix} y_0 \\ z_0 \end{pmatrix}$. The observations are $(Y(k\Delta), k = 0, \dots n)$. Since z_0 is not observed and unknown, we

 ISSN: 2102-6238

choose to add it to the parameters and set, using (S4)

$$\eta = (\alpha, z_0) \in \mathbb{R}^{a+1}; \quad \theta = (\alpha, z_0, \beta) = (\eta, \beta) \in \mathbb{R}^{a+b+1}. \tag{37}$$

The quantities introduced in (14) depend on α or η, θ and can be written $x(\eta,t) = {}^t(y(\eta,t), z(\eta,t))$, $g(\theta,t) = \begin{pmatrix} g_1(\theta,t) \\ g_2(\theta,t) \end{pmatrix}$, $R^\varepsilon(\theta,t) = \begin{pmatrix} R_1^\varepsilon(\theta,t) \\ R_2^\varepsilon(\theta,t) \end{pmatrix}$ and $\Phi(\eta,t,s)$.

The expansion of $X_\varepsilon(t)$ stated in (14) yields that $Y_\varepsilon(t)$ satisfies, using notations (36),

$$Y_\varepsilon(t) = y(\eta,t) + \varepsilon g_1(\theta,t) + \varepsilon^2 R_1^\varepsilon(\theta,t) \text{ with} \tag{38}$$

$$g_1(\theta,t) = \int_0^t (\Phi(\eta,t,u)\sigma(\beta,x(\eta,u)))_{11}\, dB_1(u) + \int_0^t (\Phi(\eta,t,u)\sigma(\beta,x(\eta,u)))_{12}\, dB_2(u). \tag{39}$$

Using that $\Phi(t,u) = \Phi(t,s)\Phi(s,u)$ yields another expression for $g_1(\theta,t_k)$

$$\begin{array}{ll} & g_1(\theta,t_k) = (\Phi(\eta,t_k,t_{k-1})g(\theta,t_{k-1}))_1 \\ + & \int_{t_{k-1}}^{t_k} (\Phi(\eta,t,u)\sigma(\beta,x(\eta,u)))_{11}\, dB_1(u) + (\Phi(\eta,t,u)\sigma(\beta,x(\eta,u)))_{12}\, dB_2(u). \end{array} \tag{40}$$

5.2. *Inference from high frequency sampling observations*

In this case of partial observations, consistent and asymptotically normal contrast-based estimators are built. Parameter identifiability is also discussed.

For estimating parameters, we use, instead of a filtering approach, the stochastic expansion of $X_\varepsilon(t)$, where the unobserved component $Z_\varepsilon(t)$ is substituted by its deterministic counterpart $z(\eta,t)$. For building a tractable estimation criterion, we also simplify the expression of $B_k(\alpha)$ (see (29)) by replacing $\Phi(\eta;t_k,t_{k-1})$ by its first order approximation, so that $\Phi_{11}(t_k,t_{k-1}) \simeq 1 + \Delta \frac{\partial b_1}{\partial y}(\alpha, x(\eta,t_{k-1}))$. The sample path used in (29) is now $\begin{pmatrix} Y(t) \\ z(\eta,t) \end{pmatrix}$ leading, instead of $B_k(\alpha)$, to the vector $\begin{pmatrix} A_k(\eta,Y) \\ 0 \end{pmatrix}$, where

$$A_k(\eta,Y) = Y(t_k) - y(\eta,t_k) - \left(1 + \Delta \frac{\partial b_1}{\partial y}(\alpha, x(\eta,t_{k-1}))\right)(Y(t_{k-1}) - y(\eta,t_{k-1})). \tag{41}$$

For a first approach, we consider an estimation criterion based on the conditional least squares built on the $A_k(\eta,Y)$'s (41):

$$\bar{U}_{\varepsilon,\Delta}(\eta,Y) = \frac{1}{\varepsilon^2 \Delta} \sum_{k=1}^{n} A_k(\eta,Y)^2. \tag{42}$$

The associated estimators are then defined as

$$\bar{\eta}_{\varepsilon,\Delta} = \underset{\eta \in K_a \times K_z}{argmin}\ \bar{U}_{\varepsilon,\Delta}(\eta,Y). \tag{43}$$

Note that this process could also be used for estimating η for fixed Δ and low frequency data, using $\Phi_{11}(t_k,t_{k-1})$ instead of its approximation.

Let us first study $\bar{U}_{\varepsilon,\Delta}(\eta,Y)$.

 ISSN: 2102-6238

Lemma 5.1. *Assume (S1)-(S5). Then, the process $\bar{U}_{\varepsilon,\Delta}(\eta,Y)$ defined in (42) satisfies, as ε and $\Delta \to 0$, that, under $\mathbb{P}^{\varepsilon}_{\theta_0}$,*

$$\varepsilon^2 \bar{U}_{\varepsilon,\Delta}(\eta,Y) \to J_T(\eta_0,\eta) = \int_0^T \Gamma_1(\eta_0,\eta;t)^2 dt \quad \textit{a.s., where} \tag{44}$$

$$\Gamma_1(\eta_0,\eta;t) = b_1(\alpha_0,x(\eta_0,t)) - b_1(\alpha,x(\eta,t)) - \frac{\partial b_1}{\partial y}(\alpha,x(\eta,t))(y(\eta_0,t)-y(\eta,t)). \tag{45}$$

So, to get that $\bar{U}_{\varepsilon,\Delta}(\eta,Y)$ is a contrast process, we need an assumption that ensures that $\{\eta \neq \eta_0 \Rightarrow J_T(\eta_0,\eta) > 0\}$. This leads to the additional identifiability assumption using (45),

(S8) : $\eta \neq \eta_0 \Rightarrow \{t \to \Gamma_1(\eta_0,\eta;t) \not\equiv 0\}$.

For deterministic systems, the notion of observability is used in the case of partial observations (Pohjanpalo, 1978; Sedoglavic, 2002), which is here $\{\eta \neq \eta_0 \Rightarrow x(\eta,.) \not\equiv x(\eta_0,.)\}$. If the underlying deterministic system is not observable, assumption (S8) which makes reference to the identifiability of the model with respect to the parameters is not satisfied. But the converse is not true, assumption (S8) being a bit stronger.

The proof of lemma 5.1 relies on two properties. First, an application of the stochastic Taylor expansion yields that, as $\varepsilon \to 0, (Y(t), 0 \le t \le T) \to (y(\eta_0,t), 0 \le t \le T)$ almost surely under $\mathbb{P}^{\varepsilon}_{\theta_0}$. Second, letting $\Delta \to 0$, we get that, uniformly with respect to $\eta = (\alpha, z_0)$,

$$sup_{k=1,\dots n} \| \frac{A_k(\eta, y(\eta_0,.))}{\Delta} - \Gamma_1(\eta_0,\eta;t_{k-1}) \| \to 0.$$

From now on, we use the convention that $\nabla_{a+1} f = \frac{\partial f}{\partial z_0}$. To study the asymptotic behaviour of $\bar{\eta}_{\varepsilon,\Delta}$, we have to introduce additional quantities. First, we define the vector $D(\eta,t) \in \mathbb{R}^{a+1}$,

$$\begin{array}{rrl} \text{if } i = 1,\dots,a, & D_i(\eta,t) = & -\frac{\partial b_1}{\partial \alpha_i}(\alpha,x(\eta,t)) - \frac{\partial b_1}{\partial z}(\alpha,x(\eta,t))\frac{\partial z}{\partial \alpha_i}(\eta,t) \\ \text{if } i = a+1, & D_i(t) = & -\frac{\partial b_1}{\partial z}(\alpha,x(\eta,t))\frac{\partial z}{\partial z_0}(\eta,t). \end{array} \tag{46}$$

Then, built on the the D_i's, define the matrix $\Lambda(\eta) = (\Lambda_{ij}(\eta))$ by

$$\Lambda_{ij}(\eta) = 2\int_0^T D_i(\eta,t) D_j(\eta,t)\, dt. \tag{47}$$

Finally, define the three functions

$$\begin{array}{rl} v_1(\theta;t) = & \sigma_{11}^2(\beta,x(\eta,t)) + \sigma_{12}^2(\beta,x(\eta,t)) = \Sigma_{11}(\beta,x(\eta,t)), \\ v_2(\theta;t,s) = & \sigma_{11}(\beta,x(s))\,(\Phi(t,s)\sigma(\beta,x(s)))_{21} + \sigma_{12}(\beta,x(s))\,(\Phi(t,s)\sigma(\beta,x(s)))_{22} \\ = & (\Phi(\eta;t,s)\Sigma(\beta,x(\eta,s)))_{21}\,, \\ v_3(\theta,t,s) = & \int_0^{t\wedge s} (\Phi(t,u)\sigma(x(u)))_{11}\,(\Phi(s,u)\sigma(x(u)))_{11}\,du \\ & + \int_0^{t\wedge s} (\Phi(t,u)\sigma(x(u)))_{22}\,(\Phi(s,u)\sigma(x(u)))_{22}\,du. \end{array} \tag{48}$$

Based on the above elements, we can now state the main result of this section.

 ISSN: 2102-6238

Theorem 5.1. *Assume (S1)-(S8). If moreover* $\Lambda(\eta_0)$ *defined in (47) is invertible, then, as* $\varepsilon, \Delta \to 0$,

$$\varepsilon^{-1}(\bar{\eta}_{\varepsilon,\Delta} - \eta_0) \to_{\mathscr{L}} \mathscr{N}(0, \Lambda(\eta_0)^{-1} V(\theta_0) \Lambda(\eta_0)^{-1}) \quad \textit{under } \mathbb{P}^{\varepsilon}_{\theta_0}, \tag{49}$$

where $V(\theta) = V^{(1)}(\theta) + V^{(2)}(\theta) + V^{(3)}(\theta)$ *with, using (46), (48),*

$$\begin{array}{rl} V^{(1)}_{ij}(\theta) = & \int_0^T D_i(\eta,t) D_j(\eta,t) v_1(\theta,t)\, dt, \\ V^{(2)}_{ij}(\theta) = & \int\int_{0\le s\le t\le T} D_i(\eta,s) D_j(\eta,t) \frac{\partial b_1}{\partial z}(\alpha, x(\eta,s)) v_2(\theta,t,s) ds\, dt, \\ V^{(3)}_{ij}(\theta) = & \int_0^T \int_0^T D_i(\eta,s) D_j(\eta,t) \frac{\partial b_1}{\partial z}(\alpha, x(\eta,s)) \frac{\partial b_1}{\partial z}(\alpha, x(\eta,t)) v_3(\theta,t,s) ds\, dt. \end{array} \tag{50}$$

The main difficulty of the proof lies in a precise study of $\varepsilon \nabla_i \bar{U}_{\varepsilon,\Delta}(\eta_0, Y)$, which is the sum of n terms that are no longer conditionally independent. The three terms in the matrix $V(\theta_0)$ come from this expansion. Indeed,

$$\varepsilon(\nabla_i \bar{U}_{\varepsilon,\Delta}(\eta_0, Y))_i, \to \mathscr{N}_{a+1}\big(0, V(\theta_0)\big) \quad \text{in distribution under } \mathbb{P}^{\varepsilon}_{\theta_0}. \tag{51}$$

Then, studying $\varepsilon^2 \nabla_{ij} \bar{U}_{\varepsilon,\Delta}(\eta, Y)$ yields, using (42), (46), as $\varepsilon, \Delta \to 0$,

$$\varepsilon^2 \nabla_{ij} \bar{U}_{\varepsilon,\Delta}(\eta_0, Y) \to \Lambda_{ij}(\eta_0) = 2 \int_0^T D_i(\eta_0,t) D_j(\eta_0,t) dt \quad \text{a.s. under } \mathbb{P}^{\varepsilon}_{\theta_0}. \tag{52}$$

5.3. Two-dimensional Ornstein-Uhlenbeck process: an explicit case

Let us describe our method on a partially observed two-dimensional Ornstein-Ulhenbeck process where all the computation are explicit. Let $X_\varepsilon(t) = \begin{pmatrix} Y_\varepsilon(t) \\ Z_\varepsilon(t) \end{pmatrix}$ satisfy

$$dX_\varepsilon(t) = AX_\varepsilon(t)dt + \varepsilon\sigma dB(t), \quad X_\varepsilon(0) = \begin{pmatrix} y_0 \\ z_0 \end{pmatrix} \text{ with } A = \begin{pmatrix} a & b \\ 0 & a+h \end{pmatrix}, \sigma = \sigma \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{53}$$

We assume that $h \neq 0$, $\sigma > 0$. The parameter in the drift is $\alpha = (a,b,h)$. For partial observations, we also need introducing $\eta = (a,b,h,z_0)$ and $\theta = (a,b,h,z_0,\sigma)$. The observations consist of $(Y(t_k), k = 0,\dots,n)$, where $T = n\Delta$ and $t_k = k\Delta$.
The solution of the ODE (8) applied to the drift of diffusion process (53) is

$$y(\eta,t) = (y_0 - \frac{z_0 b}{h})e^{at} + \frac{z_0 b}{h} e^{(a+h)t}; \quad z(\eta,t) = z_0 e^{(a+h)t}. \tag{54}$$

Assumption (S1)-(S7) are satisfied. Looking at the analytical expression of $y(\eta,t)$, we have that $bz_0 = \tilde{b}\tilde{z_0}$ leads to identical solutions $y(\eta;t)$. Therefore, assumption (S8) is not satisfied and it is impossible to estimate b and z_0 separately when observing one coordinate only. Moreover, this would also hold for the continuous observation case: the non identifiability is an intrinsic problem to the partial observation case.
So, we have to define a new parameter $b' = bz_0$ and associated $\eta' = (a,b',h)$. Then, it is easy to check that (S8) is now satisfied.
For computing the matrix $\Phi(t,u) = e^{(t-u)A}$, let decompose the matrix A as $A = PDP^{-1}$, with

 ISSN: 2102-6238

$P = \begin{pmatrix} 1 & b/h \\ 0 & 1 \end{pmatrix}, D = \begin{pmatrix} a & 0 \\ 0 & a+h \end{pmatrix}$.
Then, $\Phi(t,s) = \begin{pmatrix} e^{a(t-s)} & \frac{b}{h}\left(e^{(a+h)(t-s)} - e^{a(t-s)}\right) \\ 0 & e^{(a+h)(t-s)} \end{pmatrix}$. The solution of (53) is therefore
$X_\varepsilon(t) = Pe^{tD}P^{-1}X(0) + \varepsilon\sigma\int_0^t Pe^{(t-s)D}P^{-1}dB(s)$. Hence,

$$Y_\varepsilon(t) = y(\eta,t) + \varepsilon\sigma\left(\int_0^t e^{a(t-s)}dB_1(s) + \frac{b}{h}\int_0^t \left(e^{(a+h)(t-s)} - e^{a(t-s)}\right) dB_2(s)\right). \tag{55}$$

Using that "$\frac{\partial b_1}{\partial y}(\alpha, x(\eta,t)) = a$" and definition (54) yield that

$$A_k(\eta,Y) = Y(t_k) - y(\eta,t_k) - (1+a\Delta)\left(Y(t_{k-1}) - y(\eta,t_{k-1})\right). \tag{56}$$

The various quantities introduced in the previous section have a closed expression. The functions $D_i(\eta',t)$ defined in (46) write, using (52), (54)
$D_1(\eta',t) = D_a(\eta',t) = -y(\eta',t) - b't e^{(a+h)t} = -(y_0 - \frac{b'}{h})e^{at} - (\frac{b'}{h} + b't)e^{(a+h)t}$;
$D_2(\eta',t) = D_{b'}(\eta',t) = -e^{(a+h)t}$ and $D_3(\eta',t) = D_h(\eta',t) = -b't e^{(a+h)t}$.
The matrix $\Lambda(\eta')$ is $(\Lambda_{ij}(\eta'))_{1\le i,j\le 3}$, with $\Lambda_{ij}(\eta') = \int_0^T D_i(\eta',t)D_j(\eta',t)dt$.
The functions defined in (48) are
$v_1(\theta',t) = \sigma^2$; $v_2(\theta',t,s) = 0$; $v_3(\theta',t,s) = \sigma^2 b^2 \int_0^{t\wedge s} e^{(a+h)(t+s-2u)}du = \frac{\sigma^2 b^2}{2(a+h)}e^{2(a+h)(t\wedge s)}$.
Therefore, $V_{ij}(\theta) = \sigma^2\int_0^T D_i(\eta,t)D_j(\eta,t)dt + \frac{\sigma^2 b^2}{2(a+h)}\int_0^T\int_0^T D_i(\eta,t)D_j(\eta,s)e^{2(a+h)(t\wedge s)}ds\,dt$.
The estimator $\eta'_{\varepsilon,\Delta}$ defined by (43) is a consistent estimator of η'_0 and satisfies (49) with the matrices $\Lambda(\eta'_0)$ and $V(\theta'_0)$ obtained above. The asymptotic covariance matrix is therefore

$$\sigma^2\Lambda^{-1}(\eta') + \frac{\sigma^2 b^2}{2(a+h)}\Lambda^{-1}(\eta')\Big(\int_0^T\int_0^T D_i(\eta,t)D_j(\eta,s)e^{2(a+h)(t\wedge s)}dsdt\Big)_{ij}\Lambda^{-1}(\eta'). \tag{57}$$

In the case of complete discrete observations, the first term of (57) is the asymptotic variance obtained with conditional least squares. Therefore, the loss of information coming from partial observations is measured by the second term of (57) (added to the fact that only bz_0 is identifiable).

6. Assessment of estimator properties on simulated and real data

The properties of our minimum contrast estimators are assessed and compared to reference estimators. For the case of a completely observed process, this is performed based on simulated data (section 6.1), both on *SIR* and *SIRS* dynamics (models described in 2.2 and 2.3 respectively). For partial observations, the *SIR* case is explored on simulated data in section 6.2, whereas the *SIRS* dynamics are investigated on real data corresponding to influenza cases over several consecutive seasons in France in section 6.3. Point contrast estimates (*CE*), 95% theoretical confidence intervals (CI_{th}, when available) and empirical ones (CI_{emp}, built on 1000 runs) are provided for each set of parameter values.
For parameter dimension greater than two, confidence ellipsoids are projected on planes, by considering all pairs of parameters. Theoretical confidence ellipsoids are built as follows. By denoting $M(\theta_0)$ the variance-covariance matrix of the asymptotic normal distribution of estimators

 ISSN: 2102-6238

of parameters in drift term (i.e., I_b^{-1} in (32), I_Δ^{-1} in (34) and $\Lambda(\eta_0)^{-1}V(\theta_0)\Lambda(\eta_0)$ in (49), we have $\varepsilon^{-1}M(\theta_0)^{-1/2}(\hat{\theta}_{\varepsilon,\Delta}-\theta_0) \to_{\mathscr{L}} \mathscr{N}(0,I_k)$ (where $\hat{\theta}_{\varepsilon,\Delta}$ represents $\hat{\alpha}_{\varepsilon,\Delta}$ in (32) and (34), and $\bar{\eta}_{\varepsilon,\Delta}$ in (49) and where k is equal to a in (32) and (34), and to $a+1$ in (49)). Consequently we have,

$$\frac{1}{\varepsilon^2}\,{}^t(\hat{\theta}_{\varepsilon,\Delta}-\theta_0)M(\theta_0)^{-1}(\hat{\theta}_{\varepsilon,\Delta}-\theta_0) \to_{\mathscr{L}} \chi_2(k). \tag{58}$$

The matrix $M(\theta_0)^{-1}$ being positive, the quantity ${}^t(\hat{\theta}_{\varepsilon,\Delta}-\theta_0)M(\theta_0)^{-1}(\hat{\theta}_{\varepsilon,\Delta}-\theta_0)$ is the squared norm of vector $\hat{\theta}_{\varepsilon,\Delta}-\theta_0$ for the scalar product associated to $M(\theta_0)^{-1}$. If we denote by q_k^{95} the 95% quantile of the $\chi_2(k)$ distribution, the relation (58) leads to the definition of a theoretical asymptotic confidence ellipsoid in $\mathbb{R}^k$ as $\|\hat{\theta}_{\varepsilon,\Delta}-\theta_0\|^2_{M(\theta_0)^{-1}} \leq \varepsilon^2 q_k^{95}$.
Empirical confidence ellipsoids are based on the variance-covariance matrix of centered estimators (based on 1000 independent estimations), whose eigenvalues define the axes of ellipsoids.

6.1. The case of complete observations: SIR and SIRS models, simulated data

In this section, we consider the case where all the components of the epidemic process are observed. In the two epidemic models detailed below, this means that both $S_N(t), I_N(t)$ are observed on $[0,T]$ with sampling Δ, $((S_N(k\Delta),I_N(k\Delta)), k=1,\dots n)$ with $T=n\Delta$.
Simulated trajectories of epidemic dynamics by Markov jump processes are performed using the algorithm of Gillespie (1977) for the *SIR* model and the τ-leap method (Cao et al., 2005), more efficient for large populations, for the *SIRS* model. Based on these simulated data, the accuracy of our estimators is investigated with respect to the population size N, the number of observations n and some of the remaining parameters. Inference, using contrast (33) is based only on non extinct trajectories (chosen, according to a frequently used empirical criterion, such as the final epidemic size is larger than 5% of the number of initial susceptible individuals).

6.1.1. The SIR model

The parameters of interest for epidemics are considered following a reparameterisation: the basic reproduction number, $R_0=\frac{\lambda}{\gamma}$, which represents the average number of secondary cases generated by one infectious in a completely susceptible population, and the average infectious duration, $d=\frac{1}{\gamma}$. Two values were tested for $R_0=\{1.5,5\}$ and d was set to 3 (in days, an average value consistent with influenza infection). Three values for the population size $N=\{400,1000,10000\}$ and of the number of observations $n=\{5,10,40\}$ were considered, along with two values for the final time of observation, $T=\{20,40\}$ (in days). For each scenario defined by a combination of parameters, the analytical maximum likelihood estimator (*MLE*), calculated from the observation of all the jumps of the Markov process (Andersson and Britton, 2000), was taken as reference.

Effect of the parameter values $\{R_0,d\}$ and of the number of observations n
The accuracy of the *CE*s for $N=1000$ and from trajectories with weak ($R_0=5$) and strong ($R_0=1.5$) stochasticity is illustrated in Figure 3. R_0 and d are moderately correlated (ellipsoids are deviated with respect to the Ox and Oy axes). The shape of confidence ellipsoids depends on parameter values: for $R_0=5$, the CI_{th} is larger for R_0 than for d, whereas the opposite occurs for $R_0=1.5$. For $R_0=5$, all the CI_{th} are almost superimposed, which suggests that the estimation

 ISSN: 2102-6238

accuracy is not altered by the fact that not all jumps are observed. However, for $R_0 = 1.5$ the shape of ellipsoids varies with n. Point estimates for *MLE* and *CE* are very similar for different values of n, which confirms the use of the *CE*s in the case of a small number of observations available.

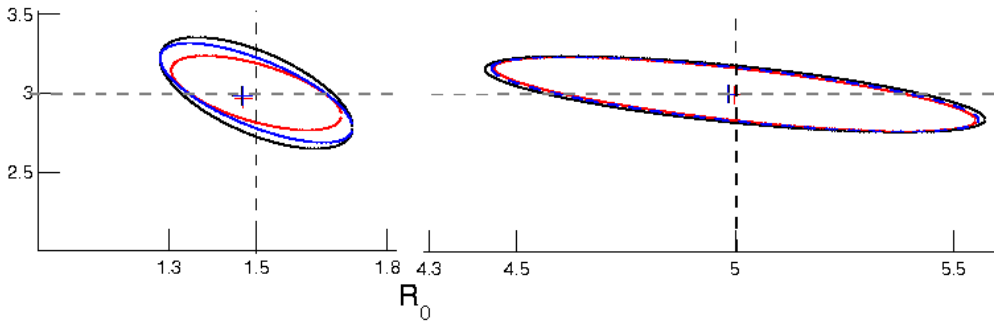

FIGURE 3. *Point estimators (+) computed by averaging over 1000 independent simulated trajectories of the SIR stochastic model completely observed and their associated theoretical confidence ellipses centered on the true value: MLE with complete observations (red), CE for one observation/day, $n = 40$ (blue) and CE for $n = 10$ (black). Two scenarios are illustrated: $(R_0, d, T) = \{(1.5, 3, 40); (5, 3, 20)\}$, with $N = 1000$. For both scenarios $(S(0), I(0)) = (0.99, 0.01)$. The value of d is reported on the y-axis. Horizontal and vertical dotted lines cross at the true value.*

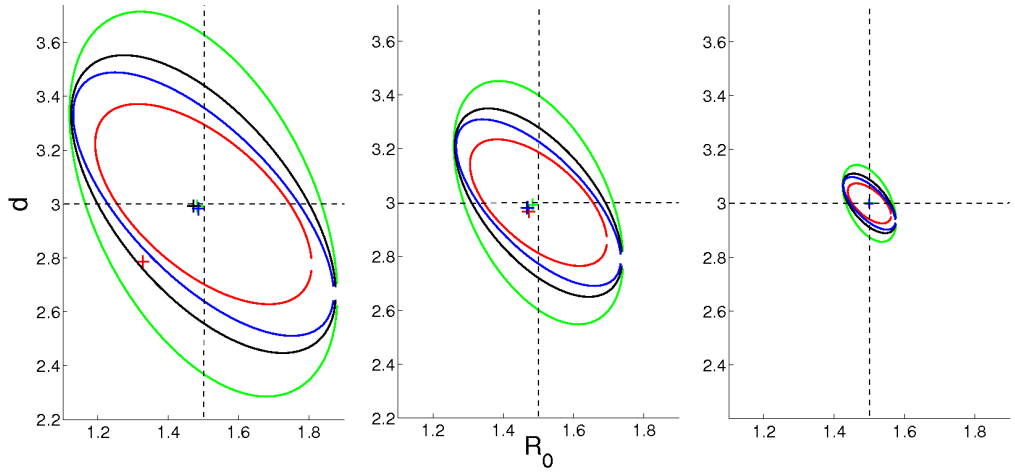

FIGURE 4. *Point estimators (+) computed by averaging over 1000 independent simulated trajectories of the SIR stochastic model completely observed and their associated theoretical confidence ellipses centered on the true value: MLE with complete observations (red), CE for one observation/day, $n = 40$ (blue), CE for $n = 10$ (black) and CE for $n = 5$ (green) for $(S(0), I(0)) = (0.99, 0.01)$, $(R_0, d) = (1.5, 3)$ and $N = \{400, 1000, 10000\}$ (from left to right). Horizontal and vertical dotted lines cross at the true value.*

Effect of the parameter values $\{R_0, d\}$ and of the population size N

From Figure 4, we can notice that $\sqrt{N}$ has an impact on estimation accuracy (the width of the confidence intervals decreases with $\sqrt{N}$). The case of very few observations ($n = 5$) leads to the largest confidence intervals. The *MLE* appears biased for $N = 400$. This could be due to the fact

 ISSN: 2102-6238

that the *MLE* is optimal when data represent a 'typical' realization (i.e. a trajectory that emerges leading to a non negligible number of infected individuals) of the Markov process, but could yield a bias when observations are far from the average behaviour. Our *CE*s seem robust to the departure from the 'typical' behaviour (i.e. for noisy trajectories obtained either for small N or small R_0).

6.1.2. *The SIRS model*

For the *SIRS* model introduced in section 2.3, four parameters were estimated: R, d, λ_1 and δ. Concerning the remaining parameters, μ was set to $1/50$ years^{-1} (a value usually considered in epidemic models), T_{per} was set to 365 days (corresponding to annual epidemics) and η was taken equal to 10^{-6} (which corresponds to 10 individuals in a population size of $N = 10^7$). We should notice that instead of estimating the real R_0 (more complicated to calculate for periodical dynamics), we prefer to estimate a parameter combination similar to the R_0 for *SIR* model, λ_0/γ, which was called here R. The performances of *CE*s were assessed on parameter combinations: $(R,d,\lambda_1,\delta) = \{(1.5,3,0.05,2)$ and $(1.5,3,0.15,2)\}$ and $T = 20$ years, with $\lambda_1 = 0.05$ leading to annual cycles and $\lambda_1 = 0.15$ to biennial dynamics (Figure 2). Numerically, the scenarios considered are consistent with influenza seasonal outbreaks. The accuracy of estimation is relatively high, as illustrated in Figure 5, regardless of the parameter. For one observation per day (which can be assimilated to a limit of data availability), the accuracy is very similar to the one based on a complete observation of the epidemic process (blue and red ellipsoids respectively). Estimations based on one observation per week are less but still reasonably accurate.

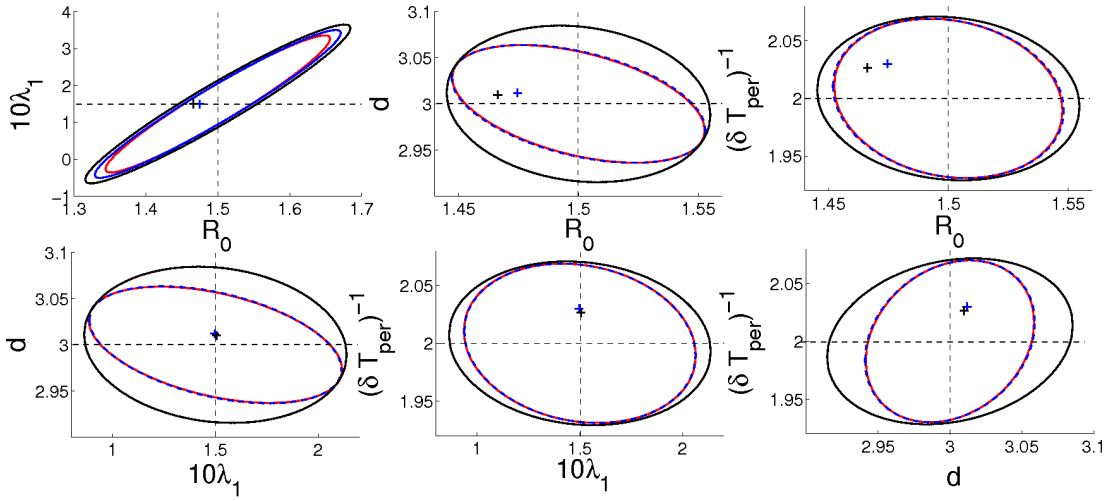

FIGURE 5. *Point estimators (+) computed by averaging over 1000 independent simulated trajectories of the SIRS stochastic model with demography and seasonal forcing in transmission, completely observed, and their associated planar projections of theoretical confidence ellipsoids centered on the true value: CE for one observation/day (blue) and for one observation/week (black) for* $(R,d,\lambda_1,\delta) = (1.5,3,0.15,2)$, $T = 20$ *years and* $N = 10^7$. *Asymptotic confidence ellipsoids* ($n \to \infty$) *are also represented (red). Horizontal and vertical dotted lines cross at the true value.*

 ISSN: 2102-6238

6.2. *The case of partial observations: SIR model, simulated data*

In this section, we consider the case where only one component of the epidemic process is observed on $[0,T]$, namely $I_N(k\Delta)$, with sampling Δ, $((S_N(k\Delta),I_N(k\Delta)),k=1,\dots n)$ with $T=n\Delta$. Theoretical identifiability of the model is first investigated and estimators based on contrast (42) along with theoretical and empirical ellipsoids are then provided and compared.

6.2.1. *Model identifiability*

To investigate the indentifiability as defined in **(S8)**, of parameters of the *SIR* ODE model, we assume that the component $i(t)$ is observed only.

Using notations from section 5.1, the drift term can be written as $b((\lambda,\gamma),(i,s)) = \begin{pmatrix} \lambda si-\gamma i \\ -\lambda si \end{pmatrix}$.

Hence, $b_1(\alpha,x(\eta,t)) = (\lambda s(t)-\gamma)i(t)$ and $\dfrac{\partial b_1}{\partial i}(\alpha,x(\eta,t)) = \lambda s(t)-\gamma$.

As a consequence, the identifiability assumption **(S8)** writes, for $\eta \neq \eta'$,

$$\Gamma_1(\eta,\eta';t) = (\lambda' s'(t)-\lambda s(t)+\gamma-\gamma')i'(t) \not\equiv 0, \tag{59}$$

where $\eta=(\lambda,\gamma,s_0)$, $\eta'=(\lambda',\gamma',s_0')$ and $x(\eta',t)=(i'(t),s'(t))$ in (45).

We assume here that the epidemic spreads, so we have $\forall t\in[0,T], i'(t)>0$ and parameters $(\lambda,\gamma,s_0,\lambda',\gamma',s_0')$ are all strictly positive. By deriving twice relation (59) we obtain the identifiability of the *SIR* parameters. Indeed, if we assume that **(S8)** is not satisfied, then, deriving (59) with respect to $s(t)$ leads to $\lambda^2 s(t)i(t)-\lambda'^2 i'(t)s'(t)=0$. We use (59) to rewrite this result as

$$\frac{s(t)}{i'(t)}(\lambda i(t)-\lambda' i'(t)) = \frac{\lambda'(\gamma'-\gamma)}{\lambda}, \tag{60}$$

$\forall t\in[0,T]$. The above term being constant, by deriving it with respect to $s(t)$ we get

$$-\lambda i(t)\frac{s(t)}{i'(t)}(\lambda i(t)-\lambda' i'(t)) = 0. \tag{61}$$

Since $i(\cdot)$ and $s(\cdot)$ are continuous and strictly positive at $t=0$, we get from (61) $\lambda i(t)=\lambda' i'(t)$, $\forall t\in[0,T]$. Morevover, since $i(0)=i'(0)$, we have that $\lambda=\lambda'$ and, from relation (60), that $\gamma=\gamma'$. This finally leads to $s_0=s_0'$, by using the two equalities obtained in (59) for $t=0$.

So, the two parameters $R_0=\lambda/\gamma$ and $d=1/\gamma$, as well as the initial state s_0 are identifiable when observing $i(t)$ only.

6.2.2. *Comparison between theoretical and empirical ellipsoids*

Performances of estimators in the case of partially observed *SIR* model are assessed on simulations obtained with the following parameters: $N=10000$, $R_0=1.5$, $d=3$, $s_0=0.97$, $T=40$. Observations are represented by vector $I_N(k\Delta)$. Estimations of parameters (R_0,d,s_0) are performed on 1000 simulated trajectories. Theoretical and empirical confidence ellipses are built as detailed in the introduction of section 6.

 ISSN: 2102-6238

In figure 6, confidence ellipsoids were truncated at plausible limits in each direction to avoid inclusion of negative values for parameters representing positive quantities and of values greater than 1 for parameters representing proportions. Quantile-based empirical 95% confidence intervals are more narrow than theoretical counterparts: $[1.38, 5.77]$ for R_0, $[2.64, 5.90]$ for d and $[0.34, 0.99]$ for s_0. The relatively unexpected large volume of confidence ellipsoids, obtained despite theoretical identifiability of model parameters when observing only one component of the system (here $I_N(k\Delta)$) is probably due to the fact that the numerical variance-covariance matrix is ill-conditioned (the order of magnitude of the third eigenvalue is 100 times smaller than that of the first two eigenvalues).

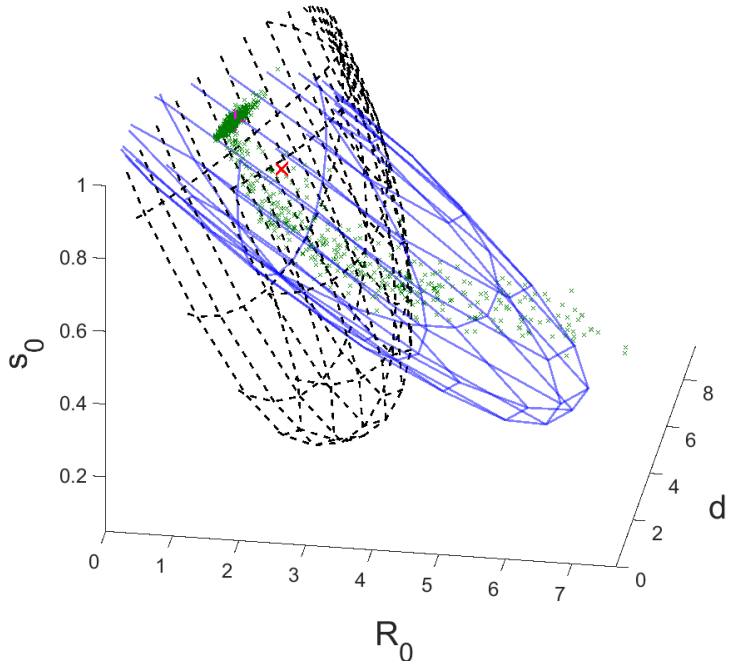

FIGURE 6. *Point estimators (green) computed by averaging over 1000 independent simulated trajectories of the SIR stochastic model, partially observed ($I(k\Delta)$ only), with true values $(R_0, d, s_0) = (1.5, 3, 0.97)$, $T = 40$ days and $N = 10000$. Theoretical confidence ellipsoid (black), centered on the true value and empirical confidence ellipsoid (blue), centered on mean estimated value are provided. Both ellipsoids are represented after truncation at plausible limits in each direction. Mean and median point estimators for (R_0, d, s_0) are $(1.89, 3.43, 0.88)$ (red* x*) and $(1.54, 3.24, 0.99)$ (purple* x*), respectively.*

6.2.3. Comparison with estimators based on complete observations

In order to assess the loss of accuracy when only partial observations are available compared to the case of complete observations, theoretical confidence ellipsoids are built for estimation from simulated data with parameter values: $N = 10000$, $R_0 = 1.5$, $d = 3$, $s_0 = 0.97$, $T = 40$. In the

 ISSN: 2102-6238

case of complete observation of the system (R_0, d) are estimated, whereas for partial observations (R_0, d, s_0) are estimated, since the initial condition (s_0, i_0) is generally not known in practice. Figure 7 illustrates these results. It shows not only the difference in accuracy between estimators in the two setups (red cross versus all other point estimates in blue), but also the effect of the number of observations on estimation accuracy, which logically increases with the number of observations.

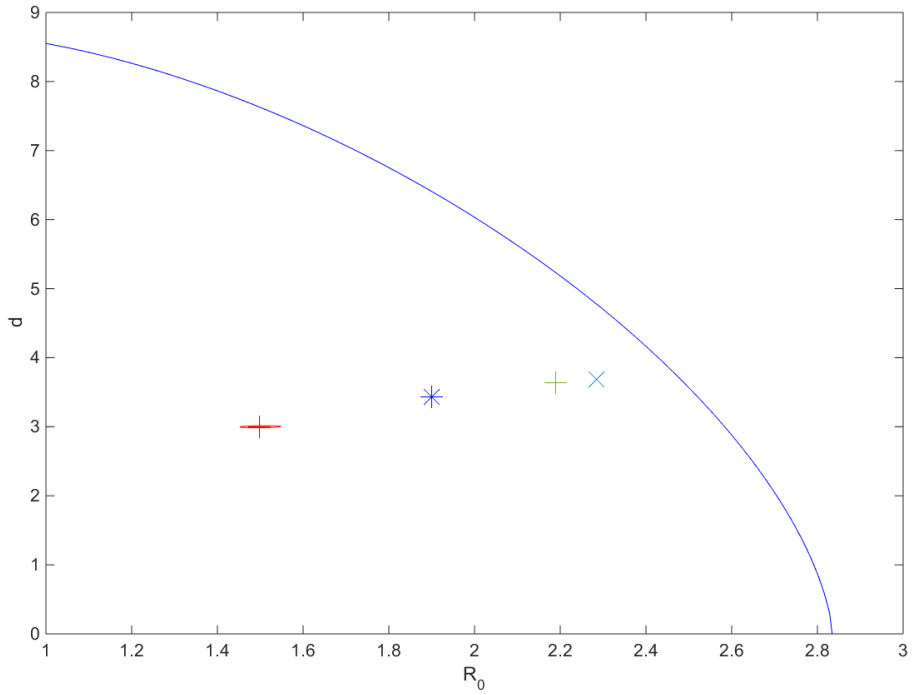

FIGURE 7. *Point estimators computed by averaging over 1000 independent simulated trajectories of the SIR stochastic model, completely observed and partially observed ($I_N(k\Delta)$ only) with true values $(R_0, d, s_0) = (1.5, 3, 0.97)$ and for $T = 40$ days, $N = 10000$ and $n \in \{1000, 100, 40\}$: complete data (red cross), and partial data with $n = 1000$ (blue star), $n = 100$ (green cross) and $n = 40$ (blue x), respectively. Theoretical confidence ellipse for complete observation (red; very narrow around the corresponding point estimator), and a planar projection (corresponding to $s_0 = 0.97$) of the theoretical ellipsoid for partial observations (blue) are also represented.*

6.3. *The case of partial observations: SIRS model, real data on influenza epidemics*

The performances of the contrast estimators for the case where only one coordinate of a diffusion process is observed are evaluated on data related to influenza outbreaks in France, collected by the French Sentinel Network (FSN), providing surveillance for several health indicators (www.sentiweb.org). These data are represented by numbers of individuals seeing a doctor during a given time interval, for symptoms related to influenza infection and are reported by a group of general practitioners (GP) voluntarily enrolled into the FSN. Several levels of errors of observation

 ISSN: 2102-6238

are associated to these data: (i) the state of individuals consulting a GP from the FSN is not exactly known: it can be assimilated to a new infection or to a new recovery, given that symptoms and infectiousness are not necessarily simultaneous and that a certain delay occurs between symptoms onset and consultation time (in fact, the observed state is more likely "infected", no more "newly infected" and not yet "newly recovered"); (ii) not all infected individuals go and see a GP; (iii) the GP's supplying the FSN database represent only a proportion of all French GP's; (iv) the exact dates of consultations are not known, data are aggregated over two-week time periods; (v) data are preprocessed by the FSN to produce observations with a daily time step.
Here, we account partly for (i) on one hand and jointly for (ii) and (iii) on the other hand and assume that observations Y_{t_k} represent a proportion of daily (observation times $t_k = k\Delta$, with $\Delta = 1$ day) numbers of newly recovered individuals: $Y_{t_k} = \rho\gamma I_{t_k}$, where ρ can be interpreted as the reporting rate. Since data are available over several seasons of influenza outbreaks (data from 1990 to 2011, hence $[0,T] = [0,21.5]$ years), an appropriate model allowing to reproduce periodic dynamics is the *SIRS* model described in section 2.3.
In summary, the data used are assumed to be discrete high frequency observations of one coordinate of the following two-dimensional diffusion with small variance:

$$\begin{cases} dS_t &= (-\lambda(t)S_t(I_t+\eta)+\delta(1-S_t-I_t)+\mu(1-S_t))dt + \frac{1}{\sqrt{N}}(\sigma_{11}dB_1(t)+\sigma_{12}dB_2(t)) \\ dI_t &= (\lambda(t)S_t(I_t+\eta)-(\gamma+\mu)I_t)dt - \frac{1}{\sqrt{N}}(\sigma_{21}dB_1(t)+\sigma_{22}dB_2(t)). \end{cases}$$

The vector of parameters to be estimated is $\alpha = (R = \lambda_0/\gamma, 10\lambda_1, d = 1/\gamma, \delta_{per} = 1/\delta T_{per}, 10\rho)$, where parameters are defined in equation (18) and more generally in the entire section 2.3. Parameters η, μ and T_{per} are fixed at plausible values: $\eta = 10^{-6}$, $\mu = \frac{1}{50}$ (years^{-1}) and $T_{per} = 365$ days. The starting point x_0 of the ODE system is unknown, but since we are interested in the stationary behavior of this process, we fix ($r_{-20T_{per}} = 0.27, i_{-20T_{per}} = 0.0001$, see Cauchemez et al., 2008 for example) and let the system evolve until $t = 0$ for the tested set of parameter α to obtain our initial starting point.

Estimation results are summarised in Figure 8, which represents multiannual dynamics of influenza cases: observed dynamics (blue curve) and simulated ones (using the ODE version of the *SIRS* model based on estimated parameter values; red curve). Estimators are associated to contrast process defined in (42). Point estimates of parameters are: $(R, 10\lambda_1, d, \delta_{per}, 10\rho) = (1.47, 1.94, 2.20, 5.66, 0.87)$. These values are in agreement with independent estimation based on data from the same database but using a different inference method, the maximum iterating filtering proposed by Bretó et al. (2009) (personal communication S. Ballesteros). As shown in Figures 6 and 7 for the *SIR* model, widths of theoretical confidence intervals for each parameter should be larger than those corresponding to complete observations of the *SIRS* model (drawn in Figure 5). In particular, for λ_1, the width of the confidence interval for partial observations will be larger than $0.35 * \sqrt{(10^7/6 * 10^7)} = 0.14$ (after correction for the population size, which is $N = 10^7$ in Figure 5 and $N = 6 * 10^7$ in Figure 8).
We can notice from Figure 8 that predicted trajectories correspond to a regime with bi-annual cycles, composed of two different peaks (red curve). The bifurcation diagram with respect to λ_1 (similar to Figure 2), when the remaining parameters are either set to fixed values (defined in this section) or to estimated values, exhibits the bifurcation from one annual cycle to bi-annual cycle

 ISSN: 2102-6238

at $\lambda_1 = 0.035$. This bifurcation value is likely to belong to the confidence interval of λ_1 (point estimate equal to 0.19 and width of the confidence interval greater than 0.14). At our knowledge, the potential influence on estimation of the proximity to a bifurcation point for models exhibiting bifurcation profiles is not well characterized.

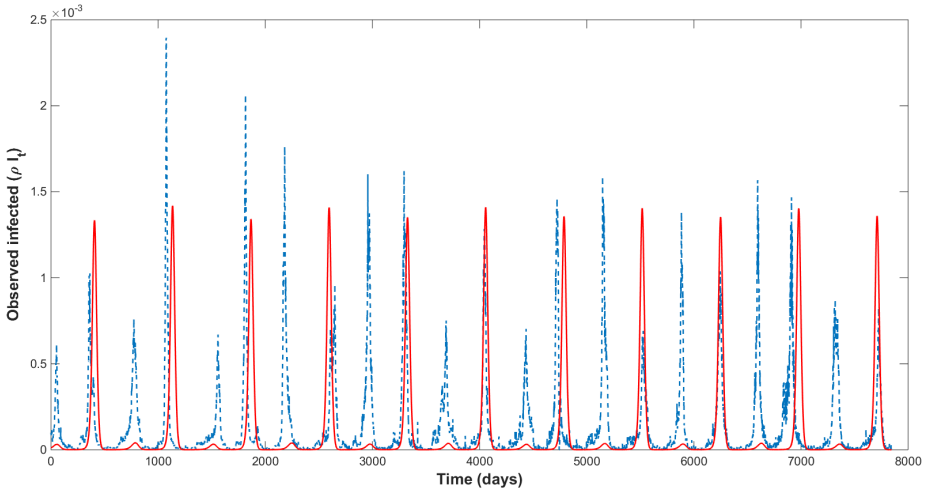

FIGURE 8. *Time series of reported cases (expressed as a fraction of the total population in France) of influenza-like illness provided by the FSN (www.sentiweb.org) (blue curve) and deterministic trajectories (mean behaviour) predicted by the SIRS model based on estimated parameters using contrast* (42) *(red curve).*

7. Discussion and concluding remarks

Several extensions to this study are possible for partial observations. First, we have chosen to detail the case of small sampling interval. The study in the case of a fixed sampling interval Δ can easily be obtained with similar tools, leading to similar results. Another extension concerns our choice of conditional least squares, by using instead an estimation criterion similar to the one used in section 4.1, substituting $S_k(\alpha,\beta)$ (see (30) or $\Sigma(\beta,X(t_k))$ by $\Sigma(\beta,x(\eta,t_k))$ for small sampling leading to the new process, using (41),

$$\bar{U}_{\varepsilon,n}(\eta,(Y(t_k))) = \sum_{k=1}^{n} log\,\Sigma(\beta,x(\eta,t_k)) + \frac{1}{\varepsilon^2\Delta}\Sigma(\beta,x(\eta,t_k))^{-1}(A_k(\eta,Y))^2. \tag{62}$$

The study of this process should yield estimators in the diffusion coefficient β, probably with additional assumptions linking ε and Δ. For fixed Δ, $S_k(\alpha,\beta)$ defined in (30) could be substituted by $(S_k(\alpha,\beta))_{11}$ in the case of distinct parameters in the drift and diffusion coefficients, and by $(S_k(\alpha,\alpha))_{11}$ in the case corresponding to epidemics where the same parameters are present in both coefficients. Finally, another extension of the method described in section 5 is the case of a p-dimensional diffusion process where only the first l-coordinates are observed (for instance the *SEIR* model with only Infected observed).

 ISSN: 2102-6238

Acknowledgements

This work was carried out with partial financial support of the French Research Agency (ANR), Program Investments for the Future; project ANR-10-BINF-07 (MIHMES). The authors thank the referees for their careful reading and helpful comments and suggestions.

References

Andersson, H. and Britton, T. (2000). *Stochastic Epidemic Models and Their Statistical Analysis*. Lecture Notes in Statistics Series. Springer.

Azencott, R. (1982). Formule de Taylor stochastique et développement asymptotique integrales de Feynmann. *Séminaire de Probabilités XVI*, pages 237–285.

Bretó, C., He, D., Ionides, E. L., and King, A. A. (2009). Time Series Analysis via Mechanistic Models. *Annals of Applied Statistics*, 3(1):319–348.

Britton, T. and Giardina, F. (2014). Introduction to statistical inference for infectious diseases. *http://arxiv.org/abs/1411.3138*.

Cao, Y., Gillespie, D. T., and Petzold, L. R. (2005). Avoiding negative populations in explicit Poisson tau-leaping. *Journal of Chemical Physics*, 123:054–104.

Cappé, O., Moulines, E., and Ryden, T. (2005). *Inference in Hidden Markov Models*. Springer Series in Statistics. Springer.

Cauchemez, S., Valleron, A.-J., Boelle, P.-Y., Flahault, A., and Ferguson, N. M. (2008). Estimating the impact of school closure on influenza transmission from Sentinel data. *Nature*, 452(7188):750–754.

Diekmann, O., Heesterbeek, H., and Britton, T. (2013). *Mathematical Tools for Understanding Infectious Disease Dynamics*. Princeton University Press.

Douc, R., Moulines, E., Olsson, J., and Van Handel, R. (2011). Consistency of the maximum likelihood estimator for general Hidden Markov Models. *Annals of Statistics*, 39(1):474–513.

Ethier, S. N. and Kurtz, T. G. (2005). *Markov Processes: Characterization and Convergence*. Wiley, 2nd edition.

Freidlin, M. and Wentzell, A. (1978). *Random Perturbations of Dynamical Systems*. Springer.

Fuchs, C. (2013). *Inference for diffusion processes*. Springer.

Genon-Catalot, V. (1990). Maximum contrast estimation for diffusion processes from discrete observations. *Statistics*, 21(1):99–116.

Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, 81(25):2340–2361.

Gloter, A. and Sørensen, M. (2009). Estimation for stochastic differential equations with a small diffusion coefficient. *Stochastic Processes and their Applications*, 119(3):679–699.

Guy, R., Larédo, C., and Vergu, E. (2014). Parametric inference for discretely observed multidimensional diffusions with small diffusion coefficient. *Stochastic Processes and their Applications*, 124:51–80.

Guy, R., Larédo, C., and Vergu, E. (2015). Approximation of epidemic models by diffusion processes and their statistical inference. *Journal of Mathematical Biology*, 70:621–646.

Jacod, J. and Shiryaev, A. N. (1987). *Limit Theorems for Stochastic Processes*. Springer.

James, M. R. and Le Gland, F. (1995). Consistent parameter estimation for partially observed diffusions with small noise. *Applied Mathematics and Optimization*, 32(1):47–72.

Karatzas, I. and Shreve, S. E. (1998). *Brownian Motion and Stochastic Calculus (Second Edition)*. Springer.

Keeling, M. J. and Rohani, P. (2011). *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press.

Kutoyants, Y. A. (1984). *Parameter estimation for stochastic processes*. Heldermann.

Kutoyants, Y. A. (1994). *Identification of Dynamical Systems with Small Noise*. Springer.

Lipster, R. S. and Shiryaev, A. N. (2001). *Statistic of Random Processes, Vol. 1*. Springer.

McKinley, T., Cook, A. R., and Deardon, R. (2009). Inference in epidemic models without likelihoods. *International Journal of Biostatistics*, 5(1).

O'Neill, P. D. (2010). Introduction and snapshot review: Relating infectious disease transmission models to data. *Statistics in medicine*, 29(20):2069–2077.

 ISSN: 2102-6238

Pohjanpalo, H. (1978). System identifiability based on the power series expansion of the solution. *Mathematical Biosciences*, 41(1-2):21–33.

Ross, J. V., Pagendam, D. E., and Polett, P. K. (2009). On parameter estimation in population models II: Multi-dimensional processes and transient dynamics. *Theoretical Population Biology*, 75(2-3):123–132.

Sedoglavic, A. (2002). A probabilistic algorithm to test local algebraic observability in polynomial time. *Journal of Symbolic Computation*, 33(5):735–755.

Sørensen, M. and Uchida, M. (2003). Small-diffusion asymptotics for discretely sampled stochastic differential equations. *Bernoulli*, 9(6):1051–1069.

van der Vaart, A. (2000). *Asymptotic statistics*. Cambridge University Press.

 ISSN: 2102-6238