



**HAL**  
open science

# A stochastic SIR model with contact-tracing: large population limits and statistical inference

Stéphan Clémenton, Viet Chi Tran, Hector de Arazoza

► **To cite this version:**

Stéphan Clémenton, Viet Chi Tran, Hector de Arazoza. A stochastic SIR model with contact-tracing: large population limits and statistical inference. 2007. hal-00171717v2

**HAL Id: hal-00171717**

**<https://hal.science/hal-00171717v2>**

Preprint submitted on 18 Feb 2008 (v2), last revised 22 Jul 2008 (v3)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A stochastic SIR model with contact-tracing: large population limits and statistical inference

Stéphan Cléménçon<sup>\*†</sup>, Viet Chi Tran<sup>‡</sup> and Hector de Arazoza<sup>§</sup>

<sup>†</sup> Laboratoire TCI UMR CNRS/Institut Telecom No.5141, FRANCE

<sup>‡</sup> Equipe Probabilités et Statistiques, Laboratoire Paul Painlevé, Université Lille 1, FRANCE

<sup>§</sup> MATCOM, Facultad de Matematica y Computacion, Universidad de la Habana, CUBA

*(Received 00 Month 200x; revised 00 Month 200x; in final form 00 Month 200x)*

This paper is devoted to present and study a specific stochastic epidemic model accounting for the effect of *contact-tracing* on the spread of an infectious disease. Precisely, one considers here the situation in which individuals identified as infected by the public health detection system may contribute to detecting other infectious individuals by providing information related to persons with whom they have had possibly infectious contacts. The control strategy, that consists in examining each individual one has been able to identify on the basis of the information collected within a certain time period, is expected to reinforce efficiently the standard random-screening based detection and slack considerably the epidemic. In the novel modelling of the spread of a communicable infectious disease considered here, the population of interest evolves through demographic, infection and detection processes, in a way that its temporal evolution is described by a stochastic Markov process, of which the component accounting for the contact-tracing feature is assumed to be valued in a space of *point measures*. For adequate scalings of the demographic, infection and detection rates, it is shown to converge to the weak deterministic solution of a PDE system, as a parameter  $n$ , interpreted as the population size roughly speaking, becomes large. From the perspective of the analysis of infectious disease data, this approximation result may serve as a key tool for exploring the asymptotic properties of standard inference methods such as maximum likelihood estimation. We state preliminary statistical results in this context. Eventually, relation of the model to the available data of the HIV epidemic in Cuba, in which country a contact-tracing detection system has been set up since 1986, is investigated and numerical applications are carried out.

*Keywords:* mathematical epidemiology, stochastic SIR model, contact-tracing, measure-valued Markov process, HIV, large population approximation, central limit theorem, maximum likelihood estimation.

*AMS Subject Classification:* 92D30, 62P10, 60F05

## 1. Introduction

Since the seminal contribution of [21], the mathematical modelling of epidemiological phenomena has received increasing attention in the applied mathematics literature. References devoted to epidemic modelling or statistical analysis of infectious disease data are much too numerous for being listed in this paper (refer to [4, 29] for recent accounts of stochastic epidemic modelling, while deterministic models for the spread of infectious diseases are comprehensively presented and discussed in [3, 9]). Here, an attempt is made to extend the 'general stochastic epidemic model', usually referred to as the standard SIR model, in order to take appropriate account of the effect of a *contact-tracing* control measure on the spread of the epidemic at the population level and to acquire a better understanding of the efficiency of this intervention strategy.

In the area of public health practice, by *contact-tracing* one means the active detection mechanism that consists in asking individuals identified as infected to name persons with whom they have had possibly infectious contacts and then, on the basis of the information provided, in striving to find those persons in order to propose them a medical examination and a cure in the event of infection. Though expensive and controversial, 'contact-tracing' programs are now receiving much attention both in the scientific literature (see [11, 15, 31] or [28] for instance) and in the public health community, within which they are generally considered as efficient guidance methods for bringing the spread of sexually transmissible diseases (STD's)

---

<sup>\*</sup>Corresponding author. Telecom Paristech Département TSI - 46, rue Barrault, 75638 Paris Cedex 13, FRANCE - tel.: +33 1 45 81 78 07- fax: +33 1 45 81 71 58 - email: stephan.clemencon@telecom-paristech.fr

under control. For instance, a contact-tracing detection system has been set up since 1986 for controlling the HIV epidemic in Cuba (refer to [10] for an overview of the evolution of HIV/AIDS in Cuba), which shall serve as a running illustration for the concepts and methods studied in the present paper. From the perspective of public health guidance practice, mathematical modelling of epidemics in presence of a contact-tracing strategy reinforcing a screening-based detection system is a crucial stake, insofar as it may help evaluating the impact of this costly control measure. In this framework, epidemic models must naturally account for the fact that, once detected, an infected person keeps on playing a role in the evolution of the epidemic for a certain time by helping towards identification of infectious individuals.

The primary goal of this paper is to generalize the standard SIR model by incorporating a *structure by age* in the subpopulation of detected individuals, age being here the time since which a person has been identified as infected. At any time, the 'R' class is described by a *point measure*, on which the contact-tracing detection rate is supposed to depend. In this manner, the way an 'R' individual contributes to contact-tracing detection may be made strongly dependent on the time since her/his detection through a given *weight function*  $\psi$ , allowing for great flexibility in the modelling. Assuming in particular a large population in which the infectious disease is spread, properties of the mathematical model are thoroughly investigated and preliminary statistical questions are tackled. Beyond stochastic modelling of the contact-tracing feature, the present work establishes large population limit results (law of large numbers and central limit theorem) for the measure-valued Markov process describing the epidemic (we follow in this respect the approach developed in [13,24,34,35]), as well as in its application to statistical analysis of the epidemic.

The paper is organized as follows. In Section 2, a Markov process with an age-structured component is introduced for modelling the temporal evolution of an epidemic in presence of contact-tracing. A short qualitative description is provided, aiming at giving an insight into how the dynamic is driven by a few key components. The process of interest is the solution of a stochastic differential equation (SDE) for which existence and uniqueness results are stated, together with a short probabilistic study. The main results of the paper are displayed in Section 3. Considering a sequence of epidemic models with contact-tracing indexed by a parameter  $n \in \mathbb{N}^*$  representing the population size roughly speaking, limit results are established when  $n \rightarrow \infty$ . Applications of the latter to the study of maximum likelihood estimators (based on complete data) in the context of statistical parametric estimation of the epidemic model with contact-tracing are then considered in Section 4. Eventually, in Section 5, these inference techniques are applied for analyzing data related to the HIV epidemic in Cuba and drawing preliminary conclusions about the effectiveness of contact-tracing in this particular case: it can be seen that the chosen model especially reflects the growing efficiency of the contact-tracing detection method, the latter becoming almost as competitive as the random screening based method ten years after the beginning of the epidemic. Technical proofs are postponed to the Appendix.

## 2. The stochastic SIR model with contact-tracing

Epidemic problems really present a great challenge to probabilists and statisticians. Models for the spread of infections are based on hypotheses about such mechanisms as infection and detection. The huge diversity of possible hypotheses could give rise to an enormous variety of probabilistic models with their specific features. Although comprehensive mathematical models should incorporate numerous features to account for real-life situations (such as population stratified according to socio-demographic characteristics, time-varying infectivity, effects of latent period, change in behavior, *etc.*), we shall deal with a stochastic epidemic model with a reasonably simple structure (a modified version of the standard 'Markovian SIR model with demography', actually), while covering some important aspects and keeping thus its pertinence from the perspective of practical applications. Indeed, incorporating too many features would naturally make the model too difficult to study analytically. As previously mentioned, we are mainly concerned here with a probabilistic modelling of the spread of an infectious disease in presence of a contact-tracing control strategy in the long-range (*i.e.* when one cannot assume that the epidemic ceases before some demographic changes occur, leading up to take into consideration immigration/birth and emigration/death processes). A parcimonious Markovian structure for describing these features is stipulated, the main novelty arising from the measure-valued component incorporated into the model in order to account for the effects of

contact-tracing. Beyond its simplicity, our modelling hopefully suffices to shed some light on the problem of investigating the efficacy of such a control measure. In the case of the HIV epidemic in Cuba for instance (see Section 5), the model obtained accounts for the fact that contact-tracing has become as efficient as random screening after 6 years. To our knowledge, earlier works have not allowed to construct a model reflecting this phenomenon (see [11] and the references therein).

### 2.1. The population dynamics

We start with a qualitative description of the population dynamics and a list of all possible events through which the population of interest may evolve (see Fig. 1). The population is structured into three classes corresponding to the different possible states with respect to the infectious disease. We adopt the standard *SIR* terminology for denoting the current status of an individual with the only differences that 'R' stands here for the population of 'removed individuals willing to take part in the contact-tracing program' and that it is structured according to the *age of detection*, namely the time since a detected individual has been identified by the public health detection system as infected. Such a distinction allows for considering heterogeneity in the way each 'R' individual contributes to the contact-tracing control. Hence, at any time  $t \geq 0$  the class of removed individuals is described by  $R_t(da)$  in  $\mathcal{M}_P(\mathbb{R}_+)$ , the set of point measures on  $\mathbb{R}_+$ : for all  $0 < a_1 < a_2 < \infty$ , the quantity  $R_t([a_1, a_2])$  represents the number of removed individuals who have been detected between times  $t - a_2$  and  $t - a_1$ . Here and throughout, we use the notation  $\langle R, \psi \rangle = \int \psi(a)R(da)$ ,  $R$  being any positive measure on  $\mathbb{R}_+$  and  $\psi$  any  $R$ -integrable function. In a more standard fashion, we shall denote by  $S_t$  and  $I_t$  the sizes of the classes of susceptible and infectious individuals.

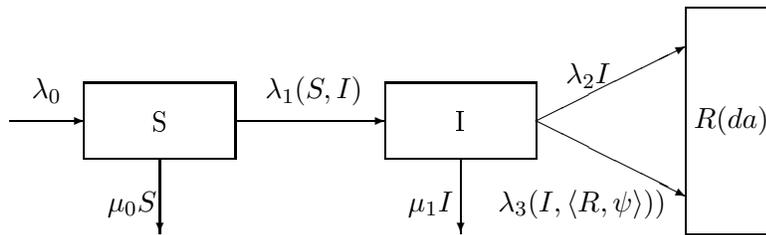


Figure 1. Flow-diagram of the *SIR* model with contact-tracing detection.

Individuals immigrate one at a time according to a Poisson process of intensity  $\lambda_0$ . Once in the population, an individual becomes 'susceptible' and may either leave the population without being contaminated (emigration or death) or independently be infected. Emigrations occur in the population at time  $t \geq 0$  with the hazard rate  $\mu_0 S_t$  and infections with the rate  $\lambda_1(S_t, I_t)$ . Once infected, an individual can be discovered by the detection system either by random screening ('spontaneous detection') or by contact-tracing, or else emigrates/dies. The hazard rates associated with these events are respectively  $\lambda_2 I_t$ ,  $\lambda_3(I_t, \langle R_t, \psi \rangle)$ , where  $\psi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is a bounded and measurable *weight function* that determines the contribution of a removed individual to the contact-tracing control according to the time  $a$  she/he has been detected (see the examples discussed below) and  $\mu_1 I_t$ . If detected, an individual takes part in the contact-tracing system by providing useful information related to her/his (possibly) infectious contacts. We do not consider the emigration/death of detected individuals since it is the availability of the information that they have given rather than their presence in the system that plays a role in the contact-tracing process. In order to avoid possible misunderstanding due to the notation, we underline that  $\lambda_1(\cdot, \cdot)$  and  $\lambda_3(\cdot, \cdot)$  here denote jump rate functions related to the *SIR* process and not the individual rates (for instance, given the class sizes  $S$  and  $I$ , the infection rate of a given susceptible is  $\lambda_1(S, I)/I$ ).

The events through which the sizes  $S_t$ ,  $I_t$  and the point measure  $R_t$  evolve are numbered as follows:

- Event  $E = 0$ : recruitment of a susceptible,
- Event  $E = 1$ : death/emigration of a susceptible,
- Event  $E = 2$ : infection,
- Event  $E = 3$ : 'spontaneous' detection of an infective,

- Event  $E = 4$ : detection of an infective by contact-tracing,
- Event  $E = 5$ : death/emigration of an infective.

Before providing a description of the population process introduced above via a system of SDE's, a few remarks and examples are in order.

**EXAMPLES. (ON MODELLING THE CONTACT-TRACING FEATURE)** *As previously explained, the removed individuals contribute to contact-tracing in function of the time since their detections and through the weight function  $\psi$ .*

1. *In the case where the information provided by a detected person enables to examine individuals at a constant rate over a period of time of fixed length  $\tau > 0$  immediately after its detection (after this time period the information is considered as consumed), the weight function could be chosen as*

$$\psi(a) = \mathbf{1}_{\{a \in [0, \tau]\}}, \text{ for all } a \geq 0.$$

*Here we have denoted by  $\mathbf{1}_{\mathcal{E}}$  the indicator function of  $\mathcal{E}$ . The second argument of the contact-tracing detection rate,  $\langle R_t, \psi \rangle$ , is then the number of individuals detected between times  $t - \tau$  and  $t$ .*

2. *The following choice:*

$$\psi(a) = e^{-c a}, \text{ for all } a \geq 0, \tag{1}$$

*with  $c > 0$ , is of particular interest when assuming that efficiency of the information provided by a detected individual (geometrically) decreases as the time  $a$  since its detection increases.*

3. *To take into account the difficulties one may encounter at the early stages of the search for contacts, we can consider functions  $\psi$  that are increasing from zero before decaying. From this viewpoint, a suitable beta or gamma density function would be possibly a reasonable choice of parametric weight function  $\psi$ .*

**Remark 1 (EXPLICIT FORMS FOR JUMP RATE FUNCTIONS)** Until now, no explicit form for the infection rate and the detection by contact-tracing rate has been specified for generality's sake (it shall be nevertheless assumed that  $\lambda_1$  and  $\lambda_3$  both fulfill the collection of assumptions **H1** listed below). In practice typical choices for the infection rate function are  $\lambda_1 SI$ ,  $\lambda_1 SI/(I + S)$  or  $\lambda_1 I$  with  $\lambda_1 > 0$  (in order to lighten notation, abusively, we shall still denote by  $\lambda_1$  and  $\lambda_3$  the parameters characterizing the parametric forms of the rate functions  $\lambda_1(\cdot, \cdot)$  and  $\lambda_3(\cdot, \cdot, \cdot)$ ). In the first example, a susceptible becomes infected with an individual rate  $\lambda_1 I$  proportional to the number of infected individuals. This rate is generally referred to as the *mass action principle* based model. In contradistinction, the two last examples correspond to a situation where the rate at which a given infective makes infectious contact does not increase with the size  $S$  of the population of susceptibles. Equivalently, the individual rate at which a susceptible becomes infected,  $\lambda_1 I/S$  or  $\lambda_1 I/(I + S)$ , decreases when  $S$  increases. Such rates are usually termed *frequency dependent*. In a similar fashion, the rate of detection by contact-tracing may be chosen as  $\lambda_3 I \langle R, \psi \rangle$ ,  $\lambda_3 I \langle R, \psi \rangle / (I + \langle R, \psi \rangle)$  or  $\lambda_3 \langle R, \psi \rangle$ , with  $\lambda_3 > 0$ .

**Remark 2 (A MORE GENERAL FRAMEWORK)** One may consider generalizations of the setup described above, stipulating for instance that the 'S' and 'I' classes are stratified according to socio-demographic features (or sexual behavior characteristics in the context of STD's) in order to account for heterogeneities caused by the social structure of the population, even if it entails introducing more duration variables in the model. One may also introduce the 'age of infection'. This would enable us to model directly time-varying infectivity, offering this way an alternative to so-called 'stage modelling' approaches (see [18] for instance). The theoretical results of this paper may be extended in a straightforward manner to such more general frameworks. In order to lighten the notation and make proofs simpler, we restrict the study to the model described above.

**Assumptions H1:** In the remainder of the paper, the rate functions  $\lambda_1$  and  $\lambda_3$  are assumed to belong to  $\mathcal{C}^1(\mathbb{R}_+^2)$ , the set of real functions of class  $\mathcal{C}^1$  on  $\mathbb{R}_+^2$ . We denote by  $\partial_S \lambda_1$ ,  $\partial_I \lambda_1$ ,  $\partial_I \lambda_3$  and  $\partial_R \lambda_3$  their partial derivatives. We also suppose that all these functions are locally Lipschitz continuous and dominated

by the mapping  $(x, x') \in \mathbb{R}_+^2 \mapsto xx'$ : for  $k \in \{1, 3\}$ , we assume that  $\forall N > 0, \exists L_k(N) > 0$  such that

$$\forall (x, x'), (y, y') \in [0, N]^2, |\lambda_k(x, x') - \lambda_k(y, y')| \leq L_k(N)(|x - y| + |x' - y'|),$$

and that  $\exists \bar{\lambda}_k > 0, \forall (x, x') \in \mathbb{R}_+^2, \lambda_k(x, x') \leq \bar{\lambda}_k xx'$  (and similarly for the partial derivatives). Finally, the weight function  $\psi$  is assumed measurable and bounded. Furthermore, for the central limit theorem, we will suppose it is of class  $\mathcal{C}^2$ .

## 2.2. On describing the epidemic by stochastic differential equations

Treading in the steps of [13] who fully developed a microscopic approach for ecological systems (see also [34, 35] where age-structure is taken into account), we now describe the temporal evolution of the epidemic by a measure-valued SDE system driven by Poisson point measures.

The process  $\{(S_t, I_t, R_t(da))\}_{t \geq 0}$  defined through the SDE system below takes its values in  $\mathbb{N} \times \mathbb{N} \times \mathcal{M}_P(\mathbb{R}_+)$  and may be seen as a generalization of the classical vector-valued Markov processes arising in stochastic SIR models.

**Definition 2.1** Consider a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , on which are defined:

- (i) a random vector  $(S_0, I_0)$  with values in  $(\mathbb{N}^*)^2$  such that  $\mathbb{E}[S_0 + I_0] < +\infty$  (at  $t = 0$ , we assume that no one has been detected yet),
- (ii) two independent Poisson point measures on  $\mathbb{R}_+^2$ ,  $Q^S(dv, du)$  and  $Q^I(dv, du)$ , with intensity  $dv \otimes du$ , the Lebesgue measure on  $\mathbb{R}_+^2$ , and independent from the initial conditions  $(S_0, I_0)$ .

Define  $\{(S_t, I_t, R_t(da))\}_{t \geq 0}$  as the Markov process solution of the following system of SDEs:

$$\begin{cases} S_t &= S_0 + \int_{v=0}^t \int_{u=0}^{\infty} (\mathbf{1}_{0 \leq u \leq \lambda_0} - \mathbf{1}_{\lambda_0 < u \leq \lambda_0 + \mu_0 S_{v-} + \lambda_1(S_{v-}, I_{v-})}) Q^S(dv, du) \\ I_t &= I_0 + \int_{v=0}^t \int_{u=0}^{\infty} \mathbf{1}_{\lambda_0 < u \leq \lambda_0 + \lambda_1(I_{v-}, S_{v-})} Q^S(dv, du) \\ &\quad - \int_{v=0}^t \int_{u=0}^{\infty} \mathbf{1}_{0 \leq u \leq (\mu_1 + \lambda_2)I_{v-} + \lambda_3(I_{v-}, \langle R_{v-}, \psi \rangle)} Q^I(dv, du) \\ \langle R_t, f \rangle &= \int_{v=0}^t \int_{u=0}^{\infty} f(0) \mathbf{1}_{0 \leq u \leq \lambda_2 I_{v-} + \lambda_3(I_{v-}, \langle R_{v-}, \psi \rangle)} Q^I(dv, du) + \int_{v=0}^t \int_{a=0}^{\infty} \partial_a f(a) R_v(da) dv, \end{cases} \quad (2)$$

for all  $f \in \mathcal{C}_b^1(\mathbb{R}_+)$  the set of real bounded functions of class  $\mathcal{C}^1$  with bounded derivatives. We have denoted by  $\partial_a f$  the gradient of  $f$  and by  $g(t-)$  the left limit in  $t \in \mathbb{R}$  of any càdlàg function  $g : \mathbb{R} \rightarrow \mathbb{R}$ .

Under **H1** and (i) of Definition 2.1, it may be seen that there exists a unique strong (non explosive) solution to SDE (2). The proof is a slight modification of the proofs of Section 2.2 in [34].

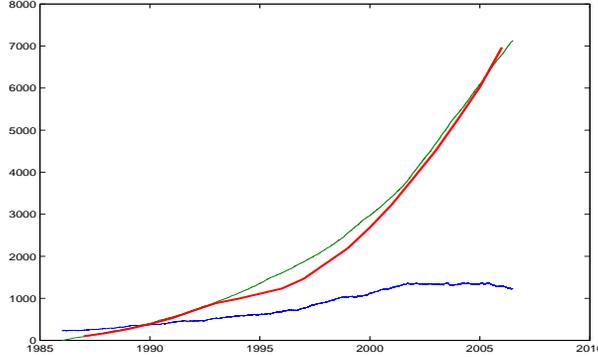


Figure 2. Simulations for the Cuban epidemic: simulated evolution of the size of the  $I$  class (blue) and of the cumulated size of the  $R$  class (green). The bold red line stands for the observed cumulated size of the  $R$  class, computed from data related to the Cuban HIV epidemic over the period 1986-2006. We have chosen  $\psi = \mathbf{1}_{[0,4]}$ . In order to mimic the change in trend that can be observed between 1995 and 2000, two periods have been separately considered. During the first period (i.e. the first fifteen years) the parameters of the simulation have been picked as follows:  $S_0 = 5 \cdot 10^6$ ,  $I_0 = 230$ ,  $\lambda_0 = 10^{-2}$ ,  $\mu_0 = 10^{-8}$ ,  $\mu_1 = 6.6 \cdot 10^{-2}$ ,  $\lambda_1 = 1.14 \cdot 10^{-7}$ ,  $\lambda_2 = 3.75 \cdot 10^{-1}$  and  $\lambda_3 = 6.55 \cdot 10^{-5}$ . In the second period, we have used:  $\lambda_1 = 1.16 \cdot 10^{-7}$ ,  $\lambda_2 = 4.45 \cdot 10^{-1}$  and  $\lambda_3 = 2.50 \cdot 10^{-4}$ .

From a practical perspective, we also emphasize that this approach paves the way for simulating trajectories of the epidemic process (see Fig. 2). An attractive advantage of stochastic models in mathematical epidemiology indeed lies in their ability to reproduce certain variability features of the observed data. As an illustration, a simulated trajectory of  $\{(I_t, \langle R_t, \mathbf{1} \rangle)\}_{t \geq 0}$  generated from Eq. (2) with jump rates  $\lambda_1 SI$  and  $\lambda_3 I \langle R, \psi \rangle$  is displayed in Fig. 2. For comparison purpose, the observed cumulative number of detected HIV+ individuals in Cuba (1986-2006) has been juxtaposed (see Section 5 for further details). It can be seen that over the first 20 years of the epidemic, the simulated and observed curves for the cumulated number of detections are fairly closed to each other, although, between years 7 and 15, the observed curve is slightly below the simulated one (it should be noticed that, during these years, less funds were available for the management of the contact-tracing detection system because of the economic crisis that followed the collapse of the Soviet Union).

### 2.3. Limiting behavior in long time asymptotics

We now state a limit result for the epidemic process introduced above, as time goes to infinity. Refer to **A1** in the Appendix for a proof based on coupling analysis.

**PROPOSITION 2.2** *Assume that  $f(a) \rightarrow 0$  as  $a \rightarrow \infty$ . Considering the Markov process  $\{(S_t, I_t, R_t(da))\}_{t \geq 0}$  introduced in Definition 2.1, we have, whatever the initial conditions  $(S_0, I_0) \in (\mathbb{N}^*)^2$ , that  $(S_t, I_t, \langle R_t(da), f \rangle) \rightarrow (S_\infty, 0, 0)$  in distribution as  $t \rightarrow \infty$ , denoting by  $S_\infty$  a Poisson random variable of parameter  $\lambda_0/\mu_0$ .*

The law of  $S_\infty$  is the stationary distribution of the  $\mathbb{N}$ -valued immigration and death process which jumps from  $k$  to  $k + 1$  with rate  $\lambda_0$  and from  $k$  to  $k - 1$  with rate  $\mu_0 k$ , and is obtained in Appendix **A1**. This ergodicity result shows that the time of extinction of the epidemic is almost surely finite, though it may be very long in practice. It is worth mentioning that in the situation of long-lasting epidemics, as in the HIV case, the long term behavior of the epidemic conditioned upon its non extinction may be refined by studying *quasi-stationary measures* (see [38] for instance). We leave this question (far from trivial when the state space is not finite) for further research.

## 3. Large population limits

The overall purpose of this section is to provide a thorough analysis of the measure-valued SIR process introduced above from the 'large population approximation' perspective (one may refer to [12] for an account of approximation theorems for Markov processes and to Chapt. 5 in [4] for applications of these concepts in approximating vector-valued SIR models), based on the recent techniques developed in [13] and [34, 35] for ecological systems.

### 3.1. Renormalization

We consider a sequence,  $\{(S_t^{(n)}, I_t^{(n)}, R_t^{(n)}(da))\}_{t \geq 0}$ ,  $n \in \mathbb{N}^*$ , of SIR processes with contact-tracing. For  $n \geq 1$ ,  $\{(S_t^{(n)}, I_t^{(n)}, R_t^{(n)}(da))\}_{t \geq 0}$  corresponds to the stochastic process described in Definition 2.1, starting from  $(S_0^{(n)}, I_0^{(n)})$  of size proportional to  $n$  and with the following rate modifications: the immigration rate is  $n\lambda_0$ , the infection jump rate function is  $n\lambda_1(S^{(n)}/n, I^{(n)}/n)$ , while the contact-tracing jump rate function is  $n\lambda_3(I^{(n)}/n, \langle R^{(n)}, \psi \rangle/n)$ . We denote by  $(s_t^{(n)}, i_t^{(n)}, r_t^{(n)}(da)) = (S_t^{(n)}/n, I_t^{(n)}/n, R_t^{(n)}(da)/n)$  the renormalized process obtained by re-weighting all individuals of the population by  $1/n$ . We assume furthermore that  $(s_0^{(n)}, i_0^{(n)})$  converges in probability to a deterministic couple  $(s_0, i_0) \in \mathbb{R}_+^{*2}$  as  $n \rightarrow \infty$ . The moment condition below shall also be required in the sequel. Let  $p > 2$ .

**Moment assumption  $M_p$ :**  $\sup_{n \in \mathbb{N}^*} \mathbb{E}[(s_0^{(n)})^p + (i_0^{(n)})^p] < +\infty$ .

This moment assumption combined with **H1** implies that the moments of order  $p$  propagate on compact time intervals  $[0, T]$  with  $T > 0$  (see [13,34] Section 3.1.2). Before writing down the martingale problem associated with  $\{(s_t^{(n)}, i_t^{(n)}, r_t^{(n)}(da))\}_{t \geq 0}$  for a given  $n \in \mathbb{N}^*$ , let us give an insight into the way the renormalizations above may be interpreted in some important examples:

*Remark 1 (ON THE MEANING OF RENORMALIZATION IN BASIC EXAMPLES)* In the case of homogeneous rate functions, the eventual impact of the renormalization on the jump rates may be described as follows.

- With  $\lambda_0^{(n)} = n\lambda_0$ , the immigration/birth rate is assumed proportional to the initial population size,
- If the form chosen for  $\lambda_1(S, I)$  is either  $\lambda_1 I$  or  $\lambda_1 SI/(I + S)$ , the renormalized infection rate function,  $\lambda_1 I^{(n)}$  or  $\lambda_1 S^{(n)} I^{(n)}/(I^{(n)} + S^{(n)})$ , is not affected by the scaling, while if one takes  $\lambda_1(S, I) = \lambda_1 SI$ , the renormalized rate function  $\lambda_1 S^{(n)} I^{(n)}/n$  decreases proportionately to  $1/n$ . This reflects the fact that for large scalings (corresponding to large "typical" population sizes) the risk of being contaminated by a given infectious individual is smaller than for small scalings.
- The same remark holds for the contact-tracing rate function  $\lambda_3(I, \langle R, \psi \rangle)$ .

The next proposition gives a semi-martingale decomposition for  $\{(s_t^{(n)}, i_t^{(n)}, r_t^{(n)}(da))\}_{t \geq 0}$ , which shall play a crucial role in our analysis.

**PROPOSITION 3.1** *Let  $n \in \mathbb{N}^*$ ,  $t \geq 0$  and  $f : (a, u) \mapsto f_u(a)$  a function in  $\mathcal{C}_b^1(\mathbb{R}_+^2)$ . Under **H1** and the moment condition  $M_p$  with  $p > 2$ ,*

$$\begin{pmatrix} M_t^{s,(n)} \\ M_t^{i,(n)} \\ M_t^{r,(n)}(f) \end{pmatrix} = \begin{pmatrix} s_t^{(n)} - s_0^{(n)} - \lambda_0 t + \int_{u=0}^t \{\mu_0 s_u^{(n)} + \lambda_1(s_u^{(n)}, i_u^{(n)})\} du \\ i_t^{(n)} - i_0^{(n)} - \int_{u=0}^t \{\lambda_1(s_u^{(n)}, i_u^{(n)}) - (\mu_1 + \lambda_2)i_u^{(n)} + \lambda_3(i_u^{(n)}, \langle r_u^{(n)}, \psi \rangle)\} du \\ \langle r_t^{(n)}, f \rangle - \int_{u=0}^t \{\langle r_u^{(n)}, \partial_a f_u + \partial_u f_u \rangle + f_u(0)(\lambda_2 i_u^{(n)} + \lambda_3(i_u^{(n)}, \langle r_u^{(n)}, \psi \rangle))\} du \end{pmatrix} \quad (3)$$

is a càdlàg  $L^2$ -martingale, with predictable quadratic variation given by:

$$\begin{cases} \langle M^{s,(n)} \rangle_t = \frac{1}{n} \int_{u=0}^t \lambda_0 + \{\mu_0 s_u^{(n)} + \lambda_1(s_u^{(n)}, i_u^{(n)})\} du \\ \langle M^{i,(n)} \rangle_t = \frac{1}{n} \int_{u=0}^t \{\lambda_1(s_u^{(n)}, i_u^{(n)}) + (\mu_1 + \lambda_2)i_u^{(n)} + \lambda_3(i_u^{(n)}, \langle r_u^{(n)}, \psi \rangle)\} du \\ \langle M^{r,(n)}(f) \rangle_t = \frac{1}{n} \int_{u=0}^t f_u^2(0) \{\lambda_2 i_u^{(n)} + \lambda_3(i_u^{(n)}, \langle r_u^{(n)}, \psi \rangle)\} du \\ \langle M^{s,(n)}, M^{i,(n)} \rangle_t = -\frac{1}{n} \int_{u=0}^t \lambda_1(s_u^{(n)}, i_u^{(n)}) du, \quad \langle M^{s,(n)}, M^{r,(n)}(f) \rangle_t = 0 \\ \langle M^{i,(n)}, M^{r,(n)}(f) \rangle_t = -\frac{1}{n} \int_{u=0}^t f_u(0) \{\lambda_2 i_u^{(n)} + \lambda_3(i_u^{(n)}, \langle r_u^{(n)}, \psi \rangle)\} du. \end{cases}$$

This result follows from the representation given in Definition 2.1, in which the Poisson measures  $Q^S$  and  $Q^I$  are introduced. It may be established by following line by line the proof of Theorem 5.2 of [13] and Theorem 3.1.8 of [34], technical details are thus omitted.

### 3.2. Main results for the large population limit

**3.2.1. Law of large numbers.** Before stating our first limit result for the sequence of renormalized SIR processes introduced above, we make clear the topology we consider. We denote by  $\mathcal{M}_F(\mathbb{R}_+)$  the space of finite measures on  $\mathbb{R}_+$ , endowed with the metrizable weak convergence topology (see [30]). For all  $n \in \mathbb{N}^*$ , the sample paths  $\{(s_t^{(n)}, i_t^{(n)}, r_t^{(n)})\}_{t \geq 0}$  belong to the Skorohod space  $\mathbb{D}(\mathbb{R}_+, \mathbb{R}_+^2 \times \mathcal{M}_F(\mathbb{R}_+))$  equipped with the metrizable  $J_1$  topology (see § 2.1 in [20] for further details).

Heuristically, since the quadratic variation of the martingale process displayed above is of order  $1/n$ , one obtains a deterministic limit by letting  $n$  tend to infinity. As a matter of fact, consider the system of

deterministic evolution equations, obtained by equating to zero the martingale process in Proposition 3.1:

$$\begin{cases} s_t &= s_0 + \int_{u=0}^t (\lambda_0 - \mu_0 s_u - \lambda_1(s_u, i_u)) du \\ i_t &= i_0 + \int_{u=0}^t (\lambda_1(s_u, i_u) - (\mu_1 + \lambda_2)i_u + \lambda_3(i_u, \langle r_u, \psi \rangle)) du \\ \langle r_t, f_t \rangle &= \int_{u=0}^t \left\{ \int_{a=0}^{\infty} (\partial_u f(a, u) + \partial_a f(a, u)) r_u(da) + f(0, u) (\lambda_2 i_u + \lambda_3(i_u, \langle r_u, \psi \rangle)) \right\} du \end{cases} \quad (4)$$

for all  $f \in C_b^1(\mathbb{R}_+^2)$ . The result below states that there exists a unique (smooth) solution to this deterministic system, to which the sequence  $\{(s^{(n)}, i^{(n)}, r^{(n)}(da))\}_{n \geq 1}$  converges in probability. This may be viewed as an extension to our measure-valued setup of the Law of Large Numbers stated in Theorem 5.2 of [4] for vector-valued SIR processes in large population asymptotics (see the references therein). A sketch of proof stands in Appendix **A2**.

**THEOREM 3.2 (LAW OF LARGE NUMBERS)** *Under **H1** and the moment condition  $M_p$  with  $p > 2$ , as the size parameter  $n$  tends to infinity, the sequence of processes  $\{(s^{(n)}, i^{(n)}, r^{(n)}(da))\}_{n \in \mathbb{N}^*}$  converges in probability in  $\mathbb{D}(\mathbb{R}_+, \mathbb{R}_+^2 \times \mathcal{M}_F(\mathbb{R}_+))$  to the unique solution  $\{(s_t, i_t, r_t(da))\}_{t \geq 0}$  of (4).*

- (i) for all  $t > 0$ , the measure  $r_t(da)$  is absolutely continuous with respect to the Lebesgue measure. Denoting by  $\rho_t(a)$  its density, the map  $(a, t) \mapsto \rho_t(a)$  is differentiable on the set  $\{a \leq t\}$  containing its support,
- (ii) the map  $t \mapsto (s_t, i_t)$  is of class  $C^1$ .

By virtue of the regularity properties mentioned above,  $(s_t, i_t, \rho_t)_{t \geq 0}$  also solves the following PDE system with initial conditions  $(s_0, i_0, \mathbf{0})$ ,  $\mathbf{0}$  denoting the constant function equal to zero:

$$\begin{cases} \frac{ds_t}{dt} &= \lambda_0 - \mu_0 s_t + \lambda_1(s_t, i_t) \\ \frac{di_t}{dt} &= \lambda_1(s_t, i_t) - (\mu_1 + \lambda_2)i_t - \lambda_3 \left( i_t, \int_{\mathbb{R}_+} \psi(a) \rho_t(a) da \right) \\ \frac{\partial \rho_t}{\partial t}(a) &= -\partial_a \rho_t(a) \\ \rho_t(0) &= \lambda_2 i_t + \lambda_3 \left( i_t, \int_{\mathbb{R}_+} \psi(a) \rho_t(a) da \right). \end{cases} \quad (5)$$

This PDE system may be seen as a generalization of deterministic epidemic models introduced in [11] (see also the references therein), taking into account the effects of the contact-tracing strategy and defined through a classical differential system. Besides, we point out that the increase of the time since detection ('detection aging') is translated into a transport equation (the third equation in (5)), with a boundary condition for  $a = 0$  (fourth equation in (5)). This is a well-known fact in age-structured population models (see [39] for instance). It is easy to prove that the solution is of the form  $\rho_t(a) = \rho_{t-a}(0)$ . One then recovers a *delay-differential equation system* (DDE), similar as those recently considered in epidemic modelling (see [7, 8, 22, 36, 37] and the references therein for instance). In this respect, it should be noticed that the DDE system (5) may be classically simplified when the weight function  $\psi$  is exponential as in Example 2. In this case, it can be replaced by a system of PDEs with finite dimensional variables only.

**3.2.2. Central limit theorem.** In order to refine the limit result stated in Theorem 3.2, we establish a central limit theorem (*CLT*), describing how the renormalized epidemic process  $(s^{(n)}, i^{(n)}, r^{(n)}(da))$  fluctuates around the solution of (5). This is an adaptation of the results obtained in Chapter 4 of [34] for age-structured birth and death processes. Let  $T > 0$  and consider the sequence of fluctuation processes:

$$\eta_t^{(n)} = \begin{pmatrix} \eta_t^{s,(n)} \\ \eta_t^{i,(n)} \\ \eta_t^{r,(n)}(da) \end{pmatrix} = \sqrt{n} \begin{pmatrix} s_t^{(n)} - s_t \\ i_t^{(n)} - i_t \\ r_t^{(n)}(da) - r_t(da) \end{pmatrix}, \quad (6)$$

$t \in [0, T]$ ,  $n \in \mathbb{N}^*$ , with values in  $\mathbb{R} \times \mathbb{R} \times \mathcal{M}_S(\mathbb{R})$ , where  $\mathcal{M}_S(\mathbb{R})$  denotes the space of signed measures on  $\mathbb{R}$  equipped with its Borel  $\sigma$ -field.

**Functional preliminaries.** Since  $\mathcal{M}_S(\mathbb{R})$  embedded with the weak convergence topology is not metrizable, we will in fact consider the sequence  $(\eta^{r,(n)}(da))_{n \in \mathbb{N}^*}$  as a sequence of processes with values in a well-chosen distribution space. In order to prove its tightness, we link this distribution space to certain Hilbert spaces. We are inspired by the works of Métivier [26], Méléard [24], and consider the following spaces:

**Definition 3.3** For  $\beta \in \mathbb{N}$ ,  $\gamma \in \mathbb{R}_+$ ,  $C^{\beta,\gamma}$  is the space of functions  $f$  of class  $\mathcal{C}^\beta$  such that  $\forall k \leq \beta$ ,  $|f^{(k)}(a)|/(1 + |a|^\gamma)$  vanishes as  $|a| \rightarrow +\infty$ , equipped with the norm:

$$\|f\|_{C^{\beta,\gamma}} := \sum_{k \leq \beta} \sup_{a \in \mathbb{R}} \frac{|f^{(k)}(a)|}{1 + |a|^\gamma}. \quad (7)$$

The spaces  $C^{\beta,\gamma}$  are Banach spaces and we denote by  $C^{-\beta,\gamma}$  their dual spaces.

$W_0^{\beta,\gamma}$  is the closure of the space  $\mathcal{C}_K^\infty(\mathbb{R})$  of infinitely differentiable functions  $f$  with compact support in  $\mathbb{R}$  for the norm  $\|\cdot\|_{W_0^{\beta,\gamma}}$  defined by:

$$\|f\|_{W_0^{\beta,\gamma}}^2 := \int_{\mathbb{R}} \sum_{k \leq \beta} \frac{|f^{(k)}(a)|^2}{1 + |a|^{2\gamma}} da. \quad (8)$$

The spaces  $W_0^{\beta,\gamma}$  are Hilbert spaces and we denote by  $W_0^{-\beta,\gamma}$  their dual spaces.

In the following, we will be interested in the following spaces (see [1, 24] for the continuous injections).

$$W_0^{3,0} \hookrightarrow C^{2,0} \hookrightarrow C^{1,0} \hookrightarrow W_0^{1,1} \hookrightarrow C^{0,1}, \quad \text{and} \quad C^{-0,1} \hookrightarrow W_0^{-1,1} \hookrightarrow C^{-1,0} \hookrightarrow C^{-2,0} \hookrightarrow W_0^{-3,0}. \quad (9)$$

The fluctuation processes are now viewed as taking their values in the dual space  $C^{-2,0}$  for technical reasons. The space  $C^{-2,0}$  is continuously included in  $W_0^{-3,0}$ , and it is in this space that the convergence in distribution stated in the next theorem is proved (see Appendix **A3** for a detailed sketch of the proof).

**THEOREM 3.4 (CENTRAL LIMIT THEOREM)** *Suppose that **H1** and the moment assumption  $M_p$  with  $p > 2$  are fulfilled and that  $\sup_{n \in \mathbb{N}^*} \mathbb{E} \left( |\eta_0^{s,(n)}|^2 + |\eta_0^{i,(n)}|^2 \right) < +\infty$ . Then,  $(\{\eta_t^{(n)}\}_{t \in [0, T]})_{n \geq 1}$ , as a sequence of random variables with values in  $\mathbb{D}([0, T], \mathbb{R} \times \mathbb{R} \times W_0^{-3,0})$ , converges in law to the solution of the following equation, that belongs to  $\mathcal{C}([0, T], \mathbb{R} \times \mathbb{R} \times C^{-2,0})$  and is unique in this space:  $\forall t \in [0, T]$ ,*

$$\eta_t = \eta_0 + W_t + \int_{u=0}^t \Psi((s_u, i_u, r_u), \eta_u) du, \quad (10)$$

where  $\Psi((s_u, i_u, r_u), \eta_u)$

$$= \begin{pmatrix} \mu_0 \eta_u^s + \partial_S \lambda_1(s_u, i_u) \eta_u^s + \partial_I \lambda_1(s_u, i_u) \eta_u^i \\ \partial_S \lambda_1(s_u, i_u) \eta_u^s + [\partial_I \lambda_1(s_u, i_u) + \mu_1 + \lambda_2 + \partial_I \lambda_3(i_u, \langle r_u, \psi \rangle)] \eta_u^i + \partial_R \lambda_3(I_u, \langle r_u, \psi \rangle) \langle \eta_u^r, \psi \rangle \\ \delta_0 [\lambda_2 + \partial_I \lambda_3(i_u, \langle r_u, \psi \rangle)] \eta_u^i + \delta_0 \partial_R \lambda_3(i_u, \langle r_u, \psi \rangle) \langle \eta_u^r, \psi \rangle + J_u^* \eta_u^r \end{pmatrix}$$

with  $\forall t \in [0, T]$ ,  $\forall f \in W_0^{3,0} (\hookrightarrow C^{2,0})$ ,  $\langle J_t^* \eta_t^r, f \rangle = \int_{\mathbb{R}_+} \partial_a f(a) \eta_t^r(da)$ , and where  $W = (W^s, W^i, W^r)$  is a continuous, centered, square-integrable Gaussian process of  $\mathcal{C}([0, T], \mathbb{R}^2 \times C^{-2,0})$ . For every  $t \in [0, T]$  and

all  $f \in C^{2,0}$ , the quadratic variation of  $(W_t^s, W_t^i, \langle W^r, f \rangle_t)_{t \in [0, T]}$  is given by:

$$\begin{cases} \langle W^s \rangle_t &= \int_0^t (\lambda_0 + \mu_0 s_u + \lambda_1(s_u, i_u)) du, \\ \langle W^i \rangle_t &= \int_0^t (\lambda_1(s_u, i_u) + (\mu_1 + \lambda_2)i_u + \lambda_3(i_u, \langle r_u, \psi \rangle)) du \\ \langle W^r(f) \rangle_t &= \int_0^t f_u^2(0) (\lambda_2 i_u + \lambda_3(i_u, \langle r_u, \psi \rangle)) du \\ \langle W^s, W^i \rangle_t &= - \int_0^t \lambda_1(s_u, i_u) du, \quad \langle W^s, \langle W^r, f \rangle \rangle_t = 0 \\ \langle W^i, \langle W^r, f \rangle \rangle_t &= - \int_0^t f_u(0) (\lambda_2 i_u + \lambda_3(i_u, \langle r_u, \psi \rangle)) du \end{cases}. \quad (11)$$

#### 4. Statistical inference by maximum likelihood estimation

We now turn to the problem of estimating the jump rates governing the dynamics of the epidemic, in a parametric setting. Although, generally, not all events related to the epidemic are observable in practice, in this premier work we deal with the ideal case where one dispose of *complete epidemic data* by means of *maximum likelihood estimation* (MLE). Indeed, MLE methods for complete data are of interest from a practical viewpoint, insofar as in certain situations they may be readily used after implementing adequate augmentation data procedures. They constitute, besides, the maximization step of (Monte Carlo-) EM procedures, which are extensively used for analyzing infectious disease data (see [5] for instance). However, we stress that the question of validly implementing EM-procedures in a continuous-time process setup is not without pitfalls (see [6, 32] and refer to [17] for an account of sequential variants of the MCEM algorithm, tailored for such a framework). Developing inference methods for the present model based on incomplete data (on the incidence process solely, for instance) shall be the scope of further research.

We start with some definitions. Let us fix the renormalization parameter  $n$ . We associate to the process  $(s_t^{(n)}, i_t^{(n)}, r_t^{(n)})_{t \geq 0}$  the sequence  $\{E_k^{(n)}, T_k^{(n)}\}_{k \in \mathbb{N}^*}$  where  $\{T_k^{(n)}\}_{k \in \mathbb{N}^*}$  is the sequence of successive jump times of the process, and where  $E_k^{(n)} \in \mathcal{E} = \{0, \dots, 5\}$  is the type of event occurring at time  $T_k^{(n)}$ ,  $k \geq 1$  (see Section 2.1). By convention, the time origin is  $T_0^{(n)} = 0$ . For notational simplicity only, the rates  $\mu_0, \mu_1, \lambda_0, \lambda_1$  are supposed to be known and we focus on the estimation of the detection rates  $\lambda_2$  and  $\lambda_3$  (extensions to a more general statistical framework are straightforward, in particular when estimating the infection rate is the matter). We suppose that the latter are entirely determined by a parameter  $\theta$ , taking values in a set  $\Theta \subset \mathbb{R}^d$ ,  $d \geq 1$ :  $\lambda_2 = \lambda_2(\theta)$  and  $\lambda_3(\cdot, \cdot) = \lambda_3(\cdot, \cdot, \theta)$ . We set  $\{\mathbb{P}_\theta\}_{\theta \in \Theta}$  the resulting family of probability measures on the underlying space  $(\Omega, \mathcal{A})$ . We denote by  $\tilde{\mathbb{P}}$  the probability measure on  $(\Omega, \mathcal{A})$  corresponding to the case when the  $(E_k^{(n)})$ 's are i.i.d. and uniformly distributed on  $\mathcal{E}$ , independent from the durations  $\Delta T_k^{(n)} = T_k^{(n)} - T_{k-1}^{(n)}$ ,  $k \in \mathbb{N}$ , supposed i.i.d. and exponentially distributed with mean  $1/(6n)$ .

##### 4.1. The likelihood function

Let  $T > 0$  and  $n \in \mathbb{N}^*$ . We denote by  $K_T^{(n)} = \sum_{k \geq 1} \mathbf{1}_{\{T_k^{(n)} \leq T\}}$  the total number of events occurring before time  $T$  and write the likelihood of  $\{(E_k^{(n)}, T_k^{(n)})\}_{1 \leq k \leq K_T^{(n)}}$ . The complete history of the epidemic until time  $T$  is described by the  $\sigma$ -field  $\mathcal{F}_t^{(n)} = \sigma\{s_u^{(n)}, i_u^{(n)}, r_u^{(n)}, u \leq t\}$ . With the notation above, the statistical model  $(\Omega, \mathcal{A}, \{\mathbb{P}_\theta\}_{\theta \in \Theta})$  is dominated along the filtration  $(\mathcal{F}_t^{(n)})_{t \geq 0}$  and  $\tilde{\mathbb{P}}$  is a dominating probability measure. In particular, for all  $\theta \in \Theta$ , we have on  $\mathcal{F}_T^{(n)}$  that  $\mathbb{P}_\theta = \mathcal{L}_T^{(n)}(\theta) \cdot \tilde{\mathbb{P}}$  with the likelihood:

$$\begin{aligned} \mathcal{L}_T^{(n)}(\theta) &= \exp \left( nT - \int_{u=0}^T (n\lambda_0 + \mu_0 n s_u^{(n)} + n\lambda_1(s_u^{(n)}, i_u^{(n)}) + (\mu_1 n + \lambda_2(\theta)n)i_u^{(n)} + n\lambda_3(i_u^{(n)}, \langle r_u^{(n)}, \psi \rangle, \theta)) du \right) \\ &\times \prod_{k=1}^{K_T^{(n)}} L_\theta(E_k, (s_{T_k}^{(n)}, i_{T_k}^{(n)}, r_{T_k}^{(n)})(da)), \end{aligned} \quad (12)$$

where:

$$L_\theta(E, (s, i, r(da))) = \lambda_0^{\mathbf{1}_{\{E=0\}}} (\mu_0 s)^{\mathbf{1}_{\{E=1\}}} \lambda_1(s, i)^{\mathbf{1}_{\{E=2\}}} (\lambda_2(\theta) i)^{\mathbf{1}_{\{E=3\}}} \lambda_3(i, \langle r, \psi \rangle, \theta)^{\mathbf{1}_{\{E=4\}}} (\mu_1 i)^{\mathbf{1}_{\{E=5\}}}.$$

If  $\theta^* \in \Theta$  denotes the 'true value' of the parameter, by taking the logarithm, keeping the terms depending on  $\theta$  solely and using the representation of Definition 2.1, one is lead to maximize the *log-likelihood*:

$$\begin{aligned} l_T^{(n)}(\theta) &= \int_{t=0}^T \int_{u=0}^\infty \left[ \log(\lambda_2(\theta) i_{t-}^{(n)}) \mathbf{1}_{\{0 \leq u \leq \lambda_2(\theta^*) n i_{t-}^{(n)}\}} \right. \\ &\quad \left. + \log(\lambda_3(i_{t-}^{(n)}, \langle r_{t-}^{(n)}, \psi \rangle, \theta)) \mathbf{1}_{\{\lambda_2(\theta^*) n i_{t-}^{(n)} < u \leq \lambda_2(\theta^*) n i_{t-}^{(n)} + n \lambda_3(i_{t-}^{(n)}, \langle r_{t-}^{(n)}, \psi \rangle, \theta^*)\}} \right] Q^I(dt, du) \\ &\quad - n \int_{t=0}^T \{\lambda_2(\theta) i_t^{(n)} + \lambda_3(i_t^{(n)}, \langle r_t^{(n)}, \psi \rangle, \theta)\} dt. \end{aligned} \quad (13)$$

#### 4.2. MLE consistency

Consider the ML estimator for  $T > 0$  and  $n \in \mathbb{N}^*$ :

$$\hat{\theta}_n = \arg \max_{\theta \in \Theta} l_T^{(n)}(\theta). \quad (14)$$

The following assumptions shall be required:

**Identifiability assumption H2:** The map  $\theta \in \Theta \mapsto (\lambda_2(\theta), \lambda_3(\cdot, \cdot, \theta))$  is injective.

**Regularity assumption R1:** For all  $(x, y) \in \mathbb{R}_+^{*2}$ , the maps  $\theta \in \Theta \mapsto \lambda_2(\theta)$  and  $\theta \in \Theta \mapsto \lambda_3(x, y, \theta)$ , are equicontinuous.

As shown by the result below, under the basic identifiability and regularity conditions stipulated above, ML estimators are consistent.

**THEOREM 4.1 (CONSISTENCY OF ML ESTIMATORS)** *Set  $\Phi(x) = \log(x) + 1/x - 1$ . Under Assumptions **H1**,  $M_p$  with  $p > 2$ , **H2** and **R1**, for all  $T > 0$  and any  $(\theta^*, \theta) \in \Theta^2$ , as  $n \rightarrow \infty$ , we have the following convergence in  $\mathbb{P}_{\theta^*}$ -probability,*

$$K_n(\theta, \theta^*) = \frac{1}{n} \{l_T^{(n)}(\theta^*) - l_T^{(n)}(\theta)\} \rightarrow K(\theta, \theta^*), \quad (15)$$

$$\text{where: } K(\theta, \theta^*) = \int_{t=0}^T \lambda_2(\theta^*) i_t^* \Phi\left(\frac{\lambda_2(\theta^*)}{\lambda_2(\theta)}\right) dt + \int_{t=0}^T \lambda_3(i_t^*, \langle r_t^*, \phi \rangle, \theta^*) \Phi\left(\frac{\lambda_3(i_t^*, \langle r_t^*, \phi \rangle, \theta^*)}{\lambda_3(i_t^*, \langle r_t^*, \phi \rangle, \theta)}\right) dt,$$

denoting by  $(s^*, i^*, r^*(da))$  the solution of the PDE system (5) with rate functions associated with  $\theta^*$ .

Under the further assumption that the parameter space  $\Theta$  is compact, the ML estimator (14) is consistent:

$$\lim_{n \rightarrow \infty} \hat{\theta}_n = \theta^*, \text{ in } \mathbb{P}_{\theta^*} \text{-probability.} \quad (16)$$

This result mainly relies on the Law of Large Numbers stated in Theorem 3.2 (see **A4** in the Appendix for technical details).

#### 4.3. MLE asymptotic normality

In order to refine our study of the asymptotic behavior of the ML estimator, we suppose that the stronger regularity assumption below is satisfied.

**Regularity condition R2:** For all  $(x, y) \in \mathbb{R}_+^{*2}$ , the maps  $\theta \in \Theta \mapsto \lambda_2(\theta)$  and  $\theta \in \Theta \mapsto \lambda_3(x, y, \theta)$  are twice continuously differentiable.

Let  $\mathcal{H}_{\theta g}$  denote the hessian matrix of any twice differentiable function  $\theta \in \Theta \mapsto g(\theta)$ . Observe that the *Fisher information matrix* is given by:

$$\begin{aligned} \mathcal{I}_{\theta} = - \int_{u=0}^T & \left\{ \mathcal{H}_{\theta} \lambda_2(\theta) i_u^* \left( \frac{\lambda_2(\theta^*)}{\lambda_2(\theta)} - 1 \right) - \nabla_{\theta} \lambda_2(\theta) \cdot {}^t \nabla_{\theta} \lambda_2(\theta) \frac{\lambda_2(\theta^*) i_u^*}{\lambda_2(\theta)^2} \right. \\ & + \mathcal{H}_{\theta} \lambda_3(i_u^*, \langle r_u^*, \psi \rangle, \theta) \left( \frac{\lambda_3(i_u^*, \langle r_u^*, \psi \rangle, \theta^*)}{\lambda_3(i_u^*, \langle r_u^*, \psi \rangle, \theta)} - 1 \right) \\ & \left. - \nabla_{\theta} \lambda_3(i_u^*, \langle r_u^*, \psi \rangle, \theta) \cdot {}^t \nabla_{\theta} \lambda_3(i_u^*, \langle r_u^*, \psi \rangle, \theta) \frac{\lambda_3(i_u^*, \langle r_u^*, \psi \rangle, \theta^*)}{\lambda_3(i_u^*, \langle r_u^*, \psi \rangle, \theta)^2} \right\} du. \quad (17) \end{aligned}$$

The next limit result then follows from Theorem 3.4 (see Appendix A5).

**THEOREM 4.2 (ASYMPTOTIC NORMALITY OF ML ESTIMATORS)** *Suppose that the assumptions of Theorem 4.1 are fulfilled with the additional condition R2. Then we have the following convergence in distribution under  $\mathbb{P}_{\theta^*}$ :*

$$\sqrt{n} \nabla_{\theta} l_T^{(n)}(\theta^*) \Rightarrow \mathcal{N}(0, \mathcal{I}_{\theta^*}), \text{ as } n \rightarrow \infty, \quad (18)$$

where  $\mathcal{I}_{\theta^*}$  is given by (17).

Moreover, if  $\mathcal{I}_{\theta^*}$  is invertible, then the ML estimator (14) is asymptotically normal: under  $\mathbb{P}_{\theta^*}$ , we have the convergence in distribution

$$\sqrt{n}(\hat{\theta}_n - \theta^*) \Rightarrow \mathcal{N}(0, \mathcal{I}_{\theta^*}^{-1}), \text{ as } n \rightarrow \infty. \quad (19)$$

#### 4.4. MLE on first simple examples

Let us consider the important situation where  $\lambda_3(i, \langle r, \psi \rangle)$  has a multiplicative form. We will consider the three following models:  $\forall i \in \mathbb{R}_+, \forall r \in \mathcal{M}_F(\mathbb{R}_+)$ ,

$$\text{Model (A): } \lambda_3(i, \langle r, \psi \rangle) = \lambda_3 \langle r, \psi \rangle, \quad (20)$$

$$\text{Model (B): } \lambda_3(i, \langle r, \psi \rangle) = \lambda_3 \frac{\langle r, \psi \rangle i}{\langle r, \psi \rangle + i}, \quad (21)$$

$$\text{Model (C): } \lambda_3(i, \langle r, \psi \rangle) = \lambda_3 \langle r, \psi \rangle i. \quad (22)$$

Here  $\theta = (\lambda_2, \lambda_3) \in \Theta \subset \mathbb{R}_+^{*2}$ , and the true parameter is denoted  $\theta^* = (\lambda_2^*, \lambda_3^*)$ . In this case, explicit maximum likelihood estimators can be obtained for  $\lambda_2$  and  $\lambda_3$ .

We differentiate the log-likelihood with respect to  $\lambda_2$  and  $\lambda_3$  and obtain *score processes* that we equate to zero. For Model (A) (see Eq. (20)), we have for  $n \in \mathbb{N}^*$  and  $t \in [0, T]$ ,

$$\begin{cases} \frac{\partial l_t^{(n)}}{\partial \lambda_2}(\lambda_2, \lambda_3) = \int_{v=0}^t \int_{u=0}^{\infty} \frac{1}{\lambda_2} \mathbf{1}_{0 \leq u \leq \lambda_2^* n i_{v-}^{(n)}} Q^I(dv, du) - n \int_{v=0}^t i_v^{(n)} dv \\ \frac{\partial l_t^{(n)}}{\partial \lambda_3}(\lambda_2, \lambda_3) = \int_{v=0}^t \int_{u=0}^{\infty} \frac{1}{\lambda_3} \mathbf{1}_{\lambda_2^* n i_{v-}^{(n)} < u \leq \lambda_2^* n i_{v-}^{(n)} + n \lambda_3^* \langle r_{v-}^{(n)}, \psi \rangle} Q^I(dv, du) - n \int_{v=0}^t \langle r_v^{(n)}, \psi \rangle dv, \end{cases}$$

and one gets:

$$\widehat{\lambda}_2^{(n)} = \frac{\int_{v=0}^T \int_{u=0}^{\infty} \mathbf{1}_{0 \leq u \leq n\lambda_2^* i_v^{(n)}} Q^I(dv, du)}{n \int_{v=0}^T i_v^{(n)} dv}, \quad \widehat{\lambda}_3^{(n,A)} = \frac{\int_{v=0}^T \int_{u=0}^{\infty} \mathbf{1}_{\lambda_2^* n i_{v-}^{(n)} < u \leq \lambda_2^* n i_{v-}^{(n)} + n\lambda_3^* \langle r_{v-}^{(n)}, \psi \rangle} Q^I(dv, du)}{n \int_0^T \langle r_v^{(n)}, \psi \rangle dv}.$$

Taking the expectation, one may immediately check that  $\mathbb{E}_{\theta^*}[(\widehat{\lambda}_2^{(n)}, \widehat{\lambda}_3^{(n,A)})] = (\lambda_2^*, \lambda_3^*)$  (denoting by  $\mathbb{E}_{\theta^*}[\cdot]$  the  $\mathbb{P}_{\theta^*}$ -expectation) and the ML estimator is unbiased. The Fisher information matrix may be easily explicited in this case:

$$\mathcal{I}_{\theta^*} = \begin{pmatrix} \int_0^T \frac{i_t^*}{\lambda_2^*} dt & 0 \\ 0 & \sigma_3^2 \end{pmatrix} \quad \text{with} \quad \sigma_3^2 = \frac{1}{\lambda_3^*} \int_0^T \langle r_t^*, \psi \rangle dt.$$

Now in the case where the contact-tracing detection rate is of the form (21) or (22), the MLE estimate  $\widehat{\lambda}_2^{(n)}$  and its asymptotic variance remain both unchanged, while we get for Model (B):

$$\widehat{\lambda}_3^{(n,B)} = \frac{\int_{v=0}^T \int_{u=0}^{\infty} \mathbf{1}_{\lambda_2^* n i_{v-}^{(n)} < u \leq \lambda_2^* n i_{v-}^{(n)} + n\lambda_3^* i_{v-}^{(n)} \langle r_{v-}^{(n)}, \psi \rangle / (i_{v-}^{(n)} + \langle r_{v-}^{(n)}, \psi \rangle)} Q^I(dv, du)}{n \int_0^T \frac{i_v^{(n)} \langle r_v^{(n)}, \psi \rangle}{(i_v^{(n)} + \langle r_v^{(n)}, \psi \rangle)} dv}$$

$$\sigma_3^2 = \frac{1}{\lambda_3^*} \int_0^T \frac{i_t^* \langle r_t^{(n)}, \psi \rangle}{(i_t^{(n)} + \langle r_t^{(n)}, \psi \rangle)} dt,$$

and for Model (C):

$$\widehat{\lambda}_3^{(n,C)} = \frac{\int_{v=0}^T \int_{u=0}^{\infty} \mathbf{1}_{\lambda_2^* n i_{v-}^{(n)} < u \leq \lambda_2^* n i_{v-}^{(n)} + n\lambda_3^* i_{v-}^{(n)} \langle r_{v-}^{(n)}, \psi \rangle} Q^I(dv, du)}{n \int_0^T i_v^{(n)} \langle r_v^{(n)}, \psi \rangle dv}, \quad \sigma_3^2 = \frac{1}{\lambda_3^*} \int_0^T i_t^* \langle r_t^{(n)}, \psi \rangle dt.$$

## 5. Application to HIV data related to the Cuban epidemic (1986-96)

This section is devoted to briefly present and discuss preliminary numerical results derived from the application of the statistical modelling previously described to real data. These data are related to the HIV epidemic in Cuba over the period 1986-1996. Our aim is to illustrate the practical interest of the theoretical notions considered in this paper. Owing to space limitations, the statistical issues of validating the parametric model stipulated below for such data shall be thoroughly investigated in a forthcoming paper, entirely dedicated to model checking, estimation and testing.

**The Cuban HIV epidemic.** In prospect of management and analysis of the epidemic, information related to the spread of HIV/AIDS in Cuba has started to be collected and gathered (in a now massive and well-documented data repository) since 1986, after the first HIV cases were detected (see the site of the World Health Organization [40] and refer to [10] for a detailed description of the HIV/AIDS epidemic in Cuba). Each time a person is detected as seropositive (HIV+), the following information is reported: date and way of detection, age, gender, area of residence, gender of sexual partners in the last two years. Furthermore, from the beginning of 1986, all detected persons are invited to give names and contact details of their recent sexual partners. Theses partners are then traced and a recommendation for HIV testing is made over a period of one year after the last sexual contact with the HIV-infected person (see [14] for further details on the Cuban management system of the AIDS epidemic). Based on the information provided, through the interview following HIV detection in the one hand and through the contact-tracing system in the other hand, a plausible date of infection for each individual is reported in the database by the Health authorities (here, as a first go, we shall use these approximate dates as if they were all exact).

After being detected, the person receives, either in the ambulatory system, regular counselling on living with HIV in order to prevent the risk of transmitting the retrovirus (defending in this respect the assumption that, once detected, a HIV carrier does not belong to the  $I$  population any more). Besides, it is essential to notice that one may assert that HIV spreads in Cuba by means of sexual transmission quasi-solely, due to certain distinctive sociological features (see [10]): indeed, since the epidemic began, injection drug use and blood transfusion accounted for a negligible number of infections, corresponding to very isolated cases.

**ML estimation.** A first attempt is now made to fit the three simple parametric SIR models with contact-tracing described in Section 4.4 with the weight function  $\psi(a) = \exp(-ca)$  (the age  $a$  being given in days and the constant  $c$  being a parameter that we will choose (see (1))). From the available data, the trajectory  $\{(s_t^{(n)}, i_t^{(n)}, r_t^{(n)}(da))\}_{t \in [0, T]}$  may be reconstructed over the 2400 first days after the first detection (more than 6.5 years). Using the results of Section 4.4, we obtain the numerical results of Table 1. We have chosen  $c = 10^{-3}$ ,  $c = 10^{-2}$  and  $c = 3 \cdot 10^{-4}$ , which correspond to various lifelengths of the information given by a detected patient to trace her/his infectious non detected partners.

model	parameter	estimated value	asymptotic std	log-likelihood
	$\lambda_2$	$9.57 \cdot 10^{-4}$	$4.34 \cdot 10^{-5}$	
$c = 10^{-2}$				
(A)	$\lambda_3$	$3.90 \cdot 10^{-3}$	$2.30 \cdot 10^{-4}$	-2115
(B)	$\lambda_3$	$4.50 \cdot 10^{-3}$	$2.67 \cdot 10^{-4}$	-2119
(C)	$\lambda_3$	$1.85 \cdot 10^{-5}$	$1.09 \cdot 10^{-6}$	-2117
$c = 10^{-3}$				
(A)	$\lambda_3$	$6.56 \cdot 10^{-4}$	$3.87 \cdot 10^{-5}$	-2138
(B)	$\lambda_3$	$1.30 \cdot 10^{-3}$	$7.72 \cdot 10^{-5}$	-2143
(C)	$\lambda_3$	$3.11 \cdot 10^{-6}$	$1.83 \cdot 10^{-7}$	-2140
$c = 3 \cdot 10^{-4}$				
(A)	$\lambda_3$	$4.37 \cdot 10^{-4}$	$2.58 \cdot 10^{-5}$	-2144
(B)	$\lambda_3$	$1.10 \cdot 10^{-3}$	$6.54 \cdot 10^{-5}$	-2146
(C)	$\lambda_3$	$2.07 \cdot 10^{-6}$	$1.22 \cdot 10^{-7}$	-2147

Table 1. *Estimated parameters and asymptotic standard deviations.*

Figure 3 shows the instantaneous detection rates  $t \mapsto \widehat{\lambda}_2^{(n)} i_t^{(n)}$  and  $t \mapsto \widehat{\lambda}_3^{(n)} (i_t^{(n)}, \langle r_t^{(n)}, \psi \rangle)$  for the three models (A), (B) and (C).

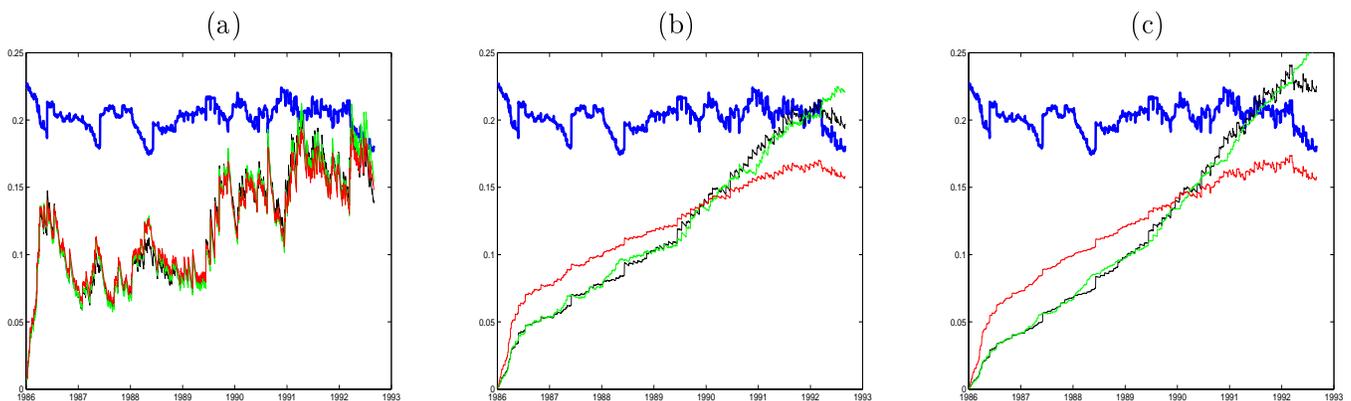


Figure 3. *Evolution of the instantaneous detection rates  $t \mapsto \widehat{\lambda}_2^{(n)} i_t^{(n)}$  (in thick blue line) and  $t \mapsto \widehat{\lambda}_3^{(n)} (i_t^{(n)}, \langle r_t^{(n)}, \psi \rangle)$  (in green (resp. red and black) line for Model (A) (resp. (B) and (C))). (a) :  $c = 10^{-2}$ , (b) :  $c = 10^{-3}$ , (c) :  $c = 3 \cdot 10^{-4}$*

Recall that  $\lambda_2$  and its asymptotic variance remain unchanged, whatever the model considered (A), (B) or (C). For the instantaneous rate of contact-tracing detection, we notice that the curves obtained by the models (A), (B) and (C) are very similar. This suggests that the model might be relatively robust to the choice of  $\psi$  (provided it has the exponential parametric form stipulated here). The curves for  $t \mapsto \widehat{\lambda}_3^{(n,A)} \langle r_t^{(n)}, \psi \rangle$  and  $t \mapsto \widehat{\lambda}_3^{(n,C)} \langle r_t^{(n)}, \psi \rangle i_t^{(n)}$  are very close. This is due to the fact that on the considered period, the number of infectious individuals remains stable. Compared with Models (A) and (C), the estimated instantaneous detection rate in Model (B) is more important in the beginning and less important in the end. This is due to the nonlinearity introduced by the denominator in (21) and to the fact that the number of detected individuals increases with time. The information given by a detected individual to trace new HIV+ ones has a smaller effect when the number of individuals is already high (certain "networks" have already been discovered then). We also underline that the larger  $c$  is and the closer the curves for the contact-tracing detection are.

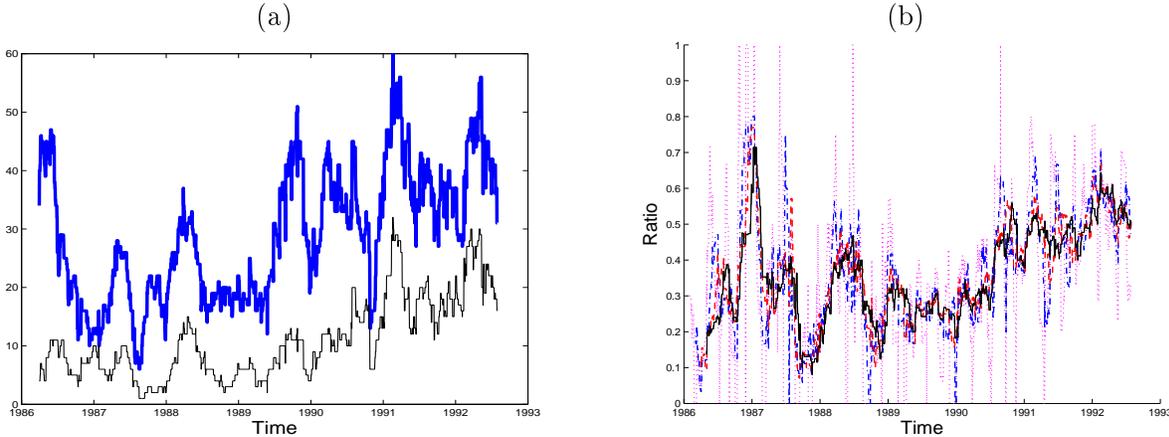


Figure 4. (a): 3-months Moving Average for the number of total detections (thin black line) and for the number of detection by contact-tracing (thick blue line). (b): Percentage of detections by contact-tracing obtained with the moving averages computed over 1 month (dotted magenta), 2 months (dash-dot blue), 3 months (dashed red) and 4 months (solid black).

On Fig. 3, it can be seen that the rate of detection by contact-tracing increases with time in each of the three models (20)-(22). It seems that, in all cases, the weight of the contact-tracing detection increases to the point of almost counterbalancing the one of the alternative way of detection. This somehow corroborates the phenomenon underlined by Fig. 4, in which the proportion of individuals detected by contact-tracing among all detected individuals computed using a moving-window of fixed length is plotted. The graphs displayed in Fig. 4 are model-free and show that the ratio of detections by contact-tracing among all detections stabilizes around 1/2, after having fluctuated during the 6 first years. These years correspond to the "burn-in period" needed for the contact-tracing detection system to start being really efficient.

Of course, a very large number of practical questions related to model fitting and interpretation/projection merit further investigation naturally arise, after this premier work. The impact of the choice of the weight function  $\psi$  on the obtained results should be carefully investigated for instance. As mentioned above, this shall be the subject of further research, much more oriented towards numerical applications and the practical use of the model.

## Acknowledgements

The authors are grateful to Dr. J. Perez of the National Institute of Tropical Diseases in Cuba for granting them access to the HIV/AIDS database. They acknowledge full support by the French Agency for Research under grant ACI-NIM no. 04-37. They also thank the referees for their judicious comments and helpful remarks.

**A1 Proof of Proposition 2.2**

Observe first that, when the Markov epidemic process  $(S, I, R(da))$  is restricted to the absorbing set  $\mathbb{N} \times \{0\} \times \mathcal{M}_P(\mathbb{R}_+)$ , the process  $(S_t)_{t \geq 0}$  evolves as an immigration and death process with immigration rate  $\lambda_0$  and death rate  $\mu_0 S_t$ . This process is classically geometrically positive recurrent. Hence, it only remains to show that the set  $\mathbb{N} \times \{0\} \times \mathcal{M}_P(\mathbb{R}_+)$  is reached in finite time with probability one, no matter what the initial state. This may be straightforwardly established by coupling analysis: it actually suffices to consider a SIR process  $(S', I', R'(da))$  with the same dynamics as  $(S, I, R(da))$  except that, when  $I = 0$  one suppose that new infectives may be recruited from the outside at a strictly positive rate. Then, one may easily see that  $(S', I')$  is an irreducible (non explosive) Markov process with state space  $\mathbb{N}^2$ . Besides, denoting by  $\Gamma$  its transition rate matrix, it is straightforward to check that the following Foster-Lyapounov's drift criterion is satisfied, with test function  $f(m, l) = m + l$ :

$$\sum_{(m', l') \in \mathbb{N}^2} \Gamma((m, l), (m', l'))(f(m', l') - f(m, l)) \leq \lambda_0 - \mu_0 m - (\mu_1 + \lambda_2)l \leq -cf(m, l) + d,$$

with  $c = \min(\mu_0, \mu_1 + \lambda_2)$  and  $d = \lambda_0$ . By virtue of Theorem 7.1 in [27],  $(S', I')$  is geometrically recurrent. Thus, when starting from  $(S_0, R_0)$ ,  $(S', I')$  and  $(S, I)$  reach  $\mathbb{N} \times \{0\}$  in a finite time  $\tau_{(S_0, R_0)}$  with finite exponential moment.

**Computation of  $S_\infty$ 's distribution.** Let us show that  $S_\infty$  is a Poisson random variable of parameter  $\lambda_0/\mu_0$ . Set  $p_k := \mathbb{P}(S_\infty = k)$  for all  $k \in \mathbb{N}$ . We have

$$\lambda_0 p_0 + \mu_0 p_1 = 0, \quad \text{and} \quad \forall k \geq 1, \lambda_0 p_{k-1} - \lambda_0 p_k + \mu_0(k+1)p_{k+1} - \mu_0 k p_k = 0. \quad (23)$$

We obtain from the first equation that  $p_1 = \lambda_0 p_0 / \mu_0$ . Assume that we have proved for  $k \in \mathbb{N}$  that:

$$\forall \ell \leq k, p_\ell = \frac{1}{\ell!} \left( \frac{\lambda_0}{\mu_0} \right)^\ell p_0. \quad (24)$$

Let us prove that (24) holds for  $k+1$ . From (23) we have:

$$p_{k+1} = \frac{p_0}{\mu_0(k+1)} \left( -\frac{\lambda_0}{(k-1)!} \left( \frac{\lambda_0}{\mu_0} \right)^{k-1} + \frac{\lambda_0}{k!} \left( \frac{\lambda_0}{\mu_0} \right)^k + \frac{\mu_0 k}{k!} \left( \frac{\lambda_0}{\mu_0} \right)^k \right) = \frac{1}{(k+1)!} \left( \frac{\lambda_0}{\mu_0} \right)^{k+1} p_0. \quad (25)$$

Since the sequence  $\{p_k\}$  defines a probability measure, we have  $\sum_{k=0}^{+\infty} p_k = 1$ . This entails that  $p_0 = e^{-\lambda_0/\mu_0}$  and the desired result then directly follows from the equation above.

**A2 Proof of Theorem 3.2 (Sketch of)**

This result may be derived from careful examination of Theorems 5.3's proof in [13] or of Theorem 3.2.2's proof in [34]. Let  $\mathcal{M}_F(\mathbb{R}_+)$  be equipped with the vague convergence topology. Applying Aldous, Rebolledo and Roelly criteria (see [2, 20, 33]), the sequence  $\{(S^{(n)}, I^{(n)}, R^{(n)})\}_{n \in \mathbb{N}^*}$  is proved tight. By Prohorov's Theorem, there hence exists a subsequence that converges in law to a limiting value  $(S, I, R) \in \mathbb{D}(\mathbb{R}_+, \mathbb{R}_+^2 \times \mathcal{M}_F(\mathbb{R}_+))$ . This subsequence can be chosen such that  $\langle R^{(n)}, 1 \rangle$  converges in law to  $\langle R, 1 \rangle$  in  $\mathbb{D}(\mathbb{R}_+, \mathbb{R}_+)$ . Since the process  $\{(S_t^{(n)}, I_t^{(n)}, R_t^{(n)})\}_{t \geq 0}$  has jumps of amplitude  $1/n$ , the limiting value is necessarily a continuous process. Using a criterion proposed in [25], one may prove that a subsequence may again be extracted from the previous one, that converges in law to  $(S, I, R)$  in  $\mathbb{D}(\mathbb{R}_+, \mathbb{R}_+^2 \times \mathcal{M}_F(\mathbb{R}_+))$ , where  $\mathcal{M}_F(\mathbb{R}_+)$

is ended with the weak convergence topology this time. The evolution equation is then identified thanks to the martingale representation provided in Proposition 3.1. The quadratic variation vanishes as  $n \rightarrow +\infty$  and the moment assumption enables us to take the limit.

### A3 Proof of Theorem 3.4 (Sketch of)

The proof of Theorem 3.4 is an adaptation of the argument developed at length in Chapter 4 of [34], which is inspired from [26] and [24]. We refer the reader to these works for hints to a complete proof and give here the main steps only. Here and throughout,  $C(T, N)$  shall denote a constant, depending on  $T$  and  $N$  only, that will not be necessarily the same at each appearance.

As previously emphasized, here and throughout the fluctuation process (6) shall be viewed as a distribution-valued process. We will deal with the spaces introduced in (9). The continuous injections which link these spaces are proved in Theorem 5.4 [1]. We consider (6) as taking its values in the space  $C^{-2,0}$ . As the image of  $C^{2,0}$  through the differential operator  $\partial_a$  is included in  $C^{1,0}$ , we will be lead to look for estimates in these both spaces. In order to work in a Hilbert setting, we consider the continuous injections linking this space with  $W_0^{-3,0}$  (in which the tightness criterion is proved) and with  $W_0^{-1,1}$  (subspace of  $C^{-1,0}$  and  $C^{-2,0}$  in which the norm of the martingale part of  $\eta^n$  is controlled). The continuous embedding  $C^{-0,1} \hookrightarrow W_0^{-1,1}$  shall also be required in order to control the norm of certain operators of  $W_0^{-1,1}$ .

**Fluctuation process.** We have the following decomposition:  $\forall f : (a, t) \mapsto f_t(a)$  in  $\mathcal{C}^1(\mathbb{R}_+^2)$ ,

$$\eta_t^{(n)}(f) = \sqrt{n} \begin{pmatrix} s_0^{(n)} - s_0 \\ i_0^{(n)} - i_0 \\ 0 \end{pmatrix} + \widetilde{M}_t^{(n)}(f) + \widetilde{V}_t^{(n)}(f), \quad \text{where:} \quad (26)$$

$$\eta_t^{(n)}(f) = \begin{pmatrix} \eta_t^{s,(n)} \\ \eta_t^{i,(n)} \\ \langle \eta_t^{r,(n)}, f_t \rangle \end{pmatrix}, \quad \widetilde{M}_t^{(n)}(f) = \begin{pmatrix} \widetilde{M}_t^{s,(n)} \\ \widetilde{M}_t^{i,(n)} \\ \widetilde{M}_t^{r,(n)}(f_t) \end{pmatrix} = \sqrt{n} \begin{pmatrix} M_t^{s,(n)} \\ M_t^{i,(n)} \\ M_t^{r,(n)}(f_t) \end{pmatrix}, \quad \widetilde{V}_t^{(n)}(f) = \begin{pmatrix} \widetilde{V}_t^{s,(n)} \\ \widetilde{V}_t^{i,(n)} \\ \widetilde{V}_t^{r,(n)}(f_t) \end{pmatrix}$$

where  $M^{s,(n)}$ ,  $M^{i,(n)}$  and  $M_t^{r,(n)}(f_t)$  have been defined in (3) and where:

$$\begin{aligned} \widetilde{V}_t^{s,(n)} &= \int_{u=0}^t \left\{ \mu_0 \eta_u^{s,(n)} + \sqrt{n}(\lambda_1(s_u^{(n)}, i_u^{(n)}) - \lambda_1(s_u, i_u)) \right\} du \\ \widetilde{V}_t^{i,(n)} &= \int_{u=0}^t \left\{ \sqrt{n}(\lambda_1(s_u^{(n)}, i_u^{(n)}) - \lambda_1(s_u, i_u)) + (\mu_1 + \lambda_2) \eta_u^{i,(n)} \right. \\ &\quad \left. + \sqrt{n}(\lambda_3(i_u^{(n)}, \langle r_u^{(n)}, \psi \rangle) - \lambda_3(i_u, \langle r_u, \psi \rangle)) \right\} du \\ \widetilde{V}_t^{r,(n)}(f_t) &= \int_{u=0}^t \left\{ f_u(0) [\lambda_2 \eta_u^{i,(n)} + \sqrt{n}(\lambda_3(i_u^{(n)}, \langle r_u^{(n)}, \psi \rangle) - \lambda_3(i_u, \langle r_u, \psi \rangle))] \right. \\ &\quad \left. + \langle \eta_u^{r,(n)}, \partial_a f_u(a) + \partial_u f_u(a) \rangle \right\} du. \end{aligned}$$

**Localization.** A difficulty arises from the fact that the size of the population is not *a priori* bounded. In this respect, for any  $N > 0$ , consider the stopping time:

$$\zeta_N^{(n)} = \inf \left\{ t \geq \mathbb{R}_+, \max \left( s_t^{(n)}, i_t^{(n)}, \langle r_t^{(n)}, 1 \rangle, \langle r_t^{(n)}, |a| \rangle \right) > N \right\}. \quad (27)$$

One may easily see that for all  $N > 0$  such that

$$N > \max\left(\sup_{t \in [0, T]} s_t, \sup_{t \in [0, T]} i_t, \sup_{t \in [0, T]} \langle r_t, 1 \rangle, \sup_{t \in [0, T]} \langle r_t, |a| \rangle\right). \quad (28)$$

we have:

$$\lim_{n \rightarrow +\infty} \mathbb{P}\left(\zeta_N^{(n)} \leq T\right) = 0 \quad (29)$$

**Moment estimates.** It is straightforward that  $\mathbb{E}[\sup_{t \in [0, T]} \|\eta_t^{r, (n)}\|_{W_0^{-1,1}}^2] \leq C(T, n)$ , for all  $n \in \mathbb{N}^*$ . Since  $C(T, n)$  depends on  $n$ , this estimate is not very interesting. Its importance lies in setting  $W_0^{-1,1}$  as a reference space. Using (9), we can then also consider it as a process with values in  $C^{-1,0}$ ,  $C^{-2,0}$  or  $W_0^{-3,0}$ . Estimates that do not depend on  $n$  can be obtained by following the proofs of Lemmas 4.4.3, 4.4.4 and 4.4.5 in [34]:

LEMMA 5.1 *Let  $N$  be fixed as in (28). Then,*

$$\begin{aligned} \sup_{n \in \mathbb{N}^*} \mathbb{E}\left[\sup_{t \in [0, T \wedge \zeta_N^{(n)}]} \{|\eta_t^{s, (n)}|^2 + |\eta_t^{i, (n)}|^2 + \|\eta_t^{r, (n)}\|_{W_0^{-3,0}}^2\}\right] \\ \leq \sup_{n \in \mathbb{N}^*} \mathbb{E}\left[\sup_{t \in [0, T \wedge \zeta_N^{(n)}]} \{|\eta_t^{s, (n)}|^2 + |\eta_t^{i, (n)}|^2 + \|\eta_t^{r, (n)}\|_{C^{-1,0}}^2\}\right] \leq C(N, T) < +\infty. \end{aligned} \quad (30)$$

*Proof* The first inequality is a consequence of the continuous injection  $C^{-1,0} \hookrightarrow W_0^{-3,0}$ . Then, (30) follows from the assumed properties of  $\lambda_1$ ,  $\lambda_3$  and  $\psi$  combined with the definition of  $\|\cdot\|_{C^{-1,0}}$  and the use of Gronwall's Lemma.  $\square$

**Tightness of the sequence**  $\{\mathcal{L}(\eta^{(n)})\}_{n \in \mathbb{N}^*}$ . By using a tightness criterion due to Métivier (see Section 2.1.5 and Theorem 2.3.2 in [20] and Corollary 1.5 and Theorem 1.6 in [26]), we will prove thanks to the preceding moment estimates that:

LEMMA 5.2 *The sequence  $\{\mathcal{L}(\eta^{(n)})\}_{n \in \mathbb{N}^*}$  of the laws of the fluctuation processes  $\{\eta^{(n)}\}_{n \in \mathbb{N}^*}$ , when considered as a sequence of  $\mathbb{D}([0, T], \mathbb{R}^2 \times W_0^{-3,0})$ , is tight.*

*Proof* Let  $\varepsilon > 0$ ,  $K > 0$ ,  $n \in \mathbb{N}^*$  and  $\phi \in W_0^{3,0}$  non zero. We start with proving that  $\sup_{t \in [0, T]} |\eta_t^{s, (n)}| + |\eta_t^{i, (n)}| + |\langle \eta_t^{r, (n)}, \phi \rangle|$  is bounded with large probability. Using the fact that, for all  $t \in [0, T]$ ,  $|\langle \eta_t^{r, (n)}, \phi \rangle| \leq \|\eta_t^{r, (n)}\|_{W_0^{-3,0}} \|\phi\|_{W_0^{3,0}} \leq \|\eta_t^{r, (n)}\|_{C^{-1,0}} \|\phi\|_{W_0^{3,0}}$ , as well as Markov's inequality, (29) and Lemma 5.1, one gets:

$$\mathbb{P}\left(\sup_{t \in [0, T]} |\eta_t^{s, (n)}| + |\eta_t^{i, (n)}| + |\langle \eta_t^{r, (n)}, \phi \rangle| > K\right) \leq \frac{C(N, T)}{K^2} + \mathbb{P}\left(\zeta_N^{(n)} \leq T\right) \leq \varepsilon, \quad (31)$$

for a proper choice of  $K$ , with  $N$  as in (28) and for sufficiently large  $n$ .

We shall now prove that the fluctuation processes 'almost belong' to a finite dimensional space. Consider a Hilbert basis  $(\varphi_k)_{k \in \mathbb{N}^*}$  of  $W_0^{3,0}$  and denote by  $\mathfrak{F}_m$  the subspace of  $W_0^{3,0}$ , with  $m \in \mathbb{N}^*$ , that is generated by  $(\varphi_k)_{k \in [1, m]}$ . Let  $\varepsilon > 0$  and  $\rho > 0$ . By Markov's inequality,

$$\mathbb{P}\left(\sup_{t \in [0, T]} \sqrt{\sum_{k=m+1}^{+\infty} \langle \eta_t^{r, (n)}, \varphi_k \rangle^2} > \rho\right) \leq \frac{1}{\rho} \mathbb{E}\left[\sup_{t \in [0, T \wedge \zeta_N^{(n)}]} \sqrt{\sum_{k=m+1}^{+\infty} \langle \eta_t^{r, (n)}, \varphi_k \rangle^2}\right] + \mathbb{P}\left(\zeta_N^{(n)} \leq T\right). \quad (32)$$

The sequence  $\left\{\sup_{t \in [0, T \wedge \zeta_N^{(n)}]} \sqrt{\sum_{k=m+1}^{+\infty} \langle \eta_t^{r, (n)}, \varphi_k \rangle^2}\right\}_{m \in \mathbb{N}^*}$  almost surely converges to 0 and is dominated

by  $\sup_{t \in [0, T \wedge \zeta_N^{(n)}]} \|\eta_t^{r, (n)}\|_{W_0^{-3,0}}$ . It is hence uniformly integrable by virtue of (30). It is thus possible to choose  $m_0 \in \mathbb{N}^*$  so that the first term in the right hand side of (32) is bounded by  $\varepsilon/2$ . The second term is bounded by  $\varepsilon/2$  using (29).

Finally, we establish an Aldous type condition for the finite-variation processes  $\{\widetilde{V}^{(n)}\}_{n \in \mathbb{N}^*}$ , the quadratic variation processes  $\{\langle \widetilde{M}^{s, (n)} \rangle\}_{n \in \mathbb{N}^*}$  and  $\{\langle \widetilde{M}^{i, (n)} \rangle\}_{n \in \mathbb{N}^*}$  and the *trace* processes  $\{\langle \widetilde{M}^{r, (n)} \rangle\}_{n \in \mathbb{N}^*}$  of  $\{\widetilde{M}^{r, (n)}\}_{n \in \mathbb{N}^*}$ , defined for a Hilbert basis  $(\varphi_k)_{k \in \mathbb{N}^*}$  of  $W_0^{1,1}$  by:  $\forall n \in \mathbb{N}^*, \forall t \in [0, T]$ ,

$$\langle \widetilde{M}^{r, (n)} \rangle_t = \int_0^t \left( \sum_{k \geq 1} \varphi_k^2(0) \left( \lambda_2 i_u^{(n)} + \lambda_3 (i_u^{(n)}, \langle r_u^{(n)}, \psi \rangle) \right) \right) du. \quad (33)$$

Since  $\sum_{k \geq 1} \varphi_k^2(0) \leq C$  (see [24]),  $\langle \widetilde{M}^{r, (n)} \rangle_t$  is  $\mathbb{P}$ -almost surely defined.

Let  $\delta > 0$  and  $(S_n, T_n)_{n \in \mathbb{N}^*}$  be a family of stopping times such that  $S_n \leq T_n \leq S_n + \delta$ . We have:  $\forall n \in \mathbb{N}^*, \forall \varsigma > 0$ ,

$$\mathbb{P} \left( \left\| \widetilde{V}_{T_n}^{r, (n)} - \widetilde{V}_{S_n}^{r, (n)} \right\|_{W_0^{-3,0}} \geq \frac{\varsigma}{3} \right) \leq \frac{9}{\varsigma^2} \mathbb{E} \left( \left\| \widetilde{V}_{T_n \wedge \zeta_N^{(n)}}^{r, (n)} - \widetilde{V}_{S_n \wedge \zeta_N^{(n)}}^{r, (n)} \right\|_{W_0^{-3,0}}^2 \right) + \mathbb{P} \left( \zeta_N^{(n)} \leq T \right). \quad (34)$$

To bound the first term, we bound  $\mathbb{E} \left[ \left\| f \mapsto \int_s^t \int_{\mathbb{R}_+} \partial_a f(a) \eta_u^{r, (n)}(da) du \right\|_{C^{-2,0}}^2 \right]$  by  $(C \int_s^t \|\eta_u^{r, (n)}\|_{C^{-1,0}} du)^2$  by noting that  $\forall f \in C^{2,0}, \partial_a f \in C^{1,0}$  with  $\|\partial_a f\|_{C^{1,0}} \leq \|f\|_{C^{2,0}}$ , and that  $\forall s, t \in [0, T], \forall n \in \mathbb{N}^*$ ,

$$\left| \int_s^t \int_{\mathbb{R}_+} \partial_a f(a) \eta_u^{r, (n)}(da) du \right| \leq \int_s^t \|\eta_u^{r, (n)}\|_{C^{-1,0}} \|\partial_a f\|_{C^{1,0}} du \leq C \int_s^t \|\eta_u^{r, (n)}\|_{C^{-1,0}} \|f\|_{C^{2,0}} du.$$

Using Lemma 5.1, we thus obtain:

$$\mathbb{P} \left( \left\| \widetilde{V}_{T_n \wedge \zeta_N^{(n)}}^{r, (n)} - \widetilde{V}_{S_n \wedge \zeta_N^{(n)}}^{r, (n)} \right\|_{W_0^{-3,0}} > \frac{\varsigma}{3} \right) < \frac{C(N)\delta^2}{\varsigma^2} + \mathbb{P} \left( \zeta_N^{(n)} \leq T \right). \quad (35)$$

Similar computations can be carried out for  $\widetilde{V}^{s, (n)}, \widetilde{V}^{i, (n)}, \langle \widetilde{M}^{s, (n)} \rangle, \langle \widetilde{M}^{i, (n)} \rangle$  and  $\langle \widetilde{M}^{r, (n)} \rangle$ .

With (31), (32) and (35), Métivier's criterion is satisfied and Lemma 5.2 is proved.  $\square$

By virtue of Prohorov's theorem, the sequence  $\{\mathcal{L}(\eta^{(n)})\}_{n \in \mathbb{N}^*}$  is relatively compact in  $\mathcal{P}(\mathbb{D}([0, T], \mathbb{R}^2 \times W_0^{-3,0}))$  embedded with the weak convergence topology. The proof of Theorem 3.4 is finished by showing that there is a unique adherence value.

**Identification of the adherence values.** Let  $\eta \in \mathbb{D}([0, T], \mathbb{R}^2 \times W_0^{-3,0})$  such that  $\mathcal{L}(\eta)$  is an adherence value of this sequence. In order to simplify notation, denote again by  $(\eta^{(n)})_{n \in \mathbb{N}^*}$  a subsequence that converges in law to  $\eta$ . Since the magnitude of the jumps of  $\eta^{(n)}$  is of order  $1/n$ , the limiting process  $\eta$  is continuous.

A first difficulty arises from the fact that Lemma 5.1 only deals with the fluctuations localized by the stopping times  $\zeta_N^{(n)}$  which depends on  $N$ .

We start off with studying the tightness and the convergence in law of the martingales  $(\widetilde{M}^{(n)})_{n \in \mathbb{N}^*}$  for which estimates that do not depend on  $n$  nor on  $N$  can be established (see [34], Lemma 4.4.5). We can prove, using the same tightness criterion as above, that  $(\widetilde{M}^{(n)})_{n \in \mathbb{N}^*}$  is tight in  $\mathbb{D}([0, T], \mathbb{R}^2 \times W_0^{-1,1})$ . Let

$W = (W^s, W^i, W^r)$  be continuous martingales as in Theorem 3.4:  $\forall \varepsilon > 0, \forall \phi \in C^{2,0}$ ,

$$\begin{aligned} & \mathbb{P}\left(\sup_{t \in [0, T]} \left| \langle \widetilde{M}^{r,(n)}(\phi) \rangle_t - \langle W^r(\phi) \rangle_t \right| > \varepsilon\right) \\ & \leq \frac{1}{\varepsilon} \mathbb{E}\left[ \sup_{t \in [0, T \wedge \zeta_N^{(n)}]} \int_0^t \left\{ \phi^2(0) \lambda_2 \frac{|\eta_s^{i,(n)}|}{\sqrt{n}} + \phi^2(0) \bar{\lambda}_3 N^2 \|\psi\|_\infty \frac{|\eta_s^{i,(n)}|}{\sqrt{n}} + L_3(N) N \frac{|\eta_s^{i,(n)}| + \|\eta_s^{r,(n)}\|_{C^{-1,0}} \|\psi\|_{C^{1,0}}}{\sqrt{n}} \right\} ds \right] \\ & \quad + \mathbb{P}\left(\zeta_N^{(n)} \leq T\right) \\ & \leq \frac{C(N, T, \|\psi\|_{C^{1,0}}) \|\phi\|_\infty}{\varepsilon \sqrt{n}} \mathbb{E}\left[ \sup_{t \in [0, T \wedge \zeta_N^{(n)}]} |\eta_t^{i,(n)}| + \|\eta_t^{r,(n)}\|_{C^{-1,0}} \right] + \mathbb{P}\left(\zeta_N^{(n)} \leq T\right). \end{aligned}$$

By (29) and by Lemma 5.1, this gives that  $(\langle \widetilde{M}^{r,(n)}(\phi) \rangle)_{n \in \mathbb{N}^*}$  converges in probability and uniformly in  $t \in [0, T]$  to  $\langle W^r(\phi) \rangle$ , defined in Theorem 3.4. Since  $\sup_{t \in [0, T]} |\Delta \widetilde{M}_t^{r,(n)}(\phi)|$  is bounded by  $C/\sqrt{n}$  and hence uniformly integrable, we obtain by applying Theorem 3.12 page 432 of Jacod and Shiryaev [19] that  $(\widetilde{M}^{r,(n)}(\phi))_{n \in \mathbb{N}^*}$  converges in law to the continuous square-integrable gaussian martingale  $W^r(\phi)$  starting from 0 and with quadratic variation given by (11). Similar computations can be done for the other terms of the bracket. Then,  $(\widetilde{M}^n)_{n \in \mathbb{N}^*}$  converges in law in  $\mathbb{D}([0, T], \mathbb{R}^2 \times W_0^{-3,0})$  to a process  $W \in \mathcal{C}([0, T], \mathbb{R}^2 \times C^{-2,0})$  such as in Theorem 3.4.

To characterize now the limit value  $\eta$ , we introduce the following functional, for  $\nu = (\nu^s, \nu^i, \nu^r) \in \mathbb{D}([0, T], \mathbb{R}^2 \times C^{-2,0})$ ,  $\phi \in C^{2,0}$  and  $t \in [0, T]$ :

$$\begin{aligned} \Psi(\nu, \phi, t) &= \begin{pmatrix} \nu_t^s \\ \nu_t^i \\ \nu_t^r(\phi) \end{pmatrix} - \begin{pmatrix} \eta_0^s \\ \eta_0^i \\ 0 \end{pmatrix} \\ & - \int_0^t \begin{pmatrix} -(\mu_0 + \partial_S \lambda_1(s_u, i_u)) \nu_u^s - \partial_I \lambda_1(s_u, i_u) \nu_u^i \\ \partial_S \lambda_1(s_u, i_u) \nu_u^s - (\partial_I \lambda_1(s_u, i_u) + \mu_1 + \lambda_2 + \partial_I \lambda_3(i_u, \langle r_u, \psi \rangle)) \nu_u^i - \partial_R \lambda_3(i_u, \langle r_u, \psi \rangle) \langle \nu_u^r, \psi \rangle \\ \phi(0) (\lambda_2 + \partial_I \lambda_3(i_u, \langle r_u, \psi \rangle)) \nu_u^i + \phi(0) \partial_R \lambda_3(i_u, \langle r_u, \psi \rangle) \langle \nu_u^r, \psi \rangle + \int_{\mathbb{R}_+} \partial_a \phi(a) \nu_u^r(da) \end{pmatrix} du \end{aligned} \quad (36)$$

Notice that, in the definition of  $\Psi$ , the density dependence has been "frozen". We can show that the process of  $\mathcal{C}([0, T], \mathbb{R}^2 \times C^{-2,0})$  defined for every  $\phi \in C^{2,0}$  and  $t \in [0, T]$  by  $(\widetilde{M}_t^s, \widetilde{M}_t^i, \widetilde{M}_t^r(\phi)) := \Psi(\eta, \phi, t)$  has the same law as the process  $W$  defined in Theorem 3.4.

Indeed, since  $(\eta^{(n)})_{n \in \mathbb{N}^*}$  converges in law to the continuous process  $\eta$ , we have  $\forall \phi \in C^{2,0}, \lim_{n \rightarrow +\infty} \Psi(\eta^{(n)}, \phi, \cdot) = \Psi(\eta, \phi, \cdot)$ . We shall now prove that  $\Psi(\eta^{(n)}, \phi, \cdot)$  has the same limit in law as  $\widetilde{M}^{(n)}(\phi)$ .

From (6) and (36):  $\forall n \in \mathbb{N}^*, \forall \phi \in C^{2,0}, \forall t \in [0, T]$ ,

$$\begin{aligned} |\Psi(\eta^{(n)}, \phi, t) - (\widetilde{M}_t^{s,(n)}, \widetilde{M}_t^{i,(n)}, \widetilde{M}_t^{r,(n)}(\phi))|^2 &= \left( \int_0^t A(n, s) ds \right)^2 + \left( \int_0^t [A(n, s) + B(n, \phi, s)] ds \right)^2 \\ & \quad + \left( \int_0^t \phi(0) B(n, \phi, s) ds \right)^2, \end{aligned} \quad (37)$$

where:

$$\begin{aligned} A(n, u) &= \partial_S \lambda_1(s_u, i_u) \eta_u^{s,(n)} + \partial_I \lambda_1(s_u, i_u) \eta_u^{i,(n)} - \sqrt{n} \{ \lambda_1(s_u^{(n)}, i_u^{(n)}) - \lambda_1(s_u, i_u) \}, \\ B(n, s) &= \{ \partial_I \lambda_3(i_u, \langle r_u, \psi \rangle) \eta_u^{i,(n)} - \partial_R \lambda_3(i_u, \langle r_u, \psi \rangle) \langle \eta_u^{r,(n)}, \psi \rangle \} - \sqrt{n} \{ \lambda_3(i_u^{(n)}, \langle r_u^{(n)}, \psi \rangle) - \lambda_3(i_u, \langle r_u, \psi \rangle) \}. \end{aligned}$$

We obtain from

$$\begin{aligned}\lambda_1(s_u^{(n)}, i_u^{(n)}) - \lambda_1(s_u, i_u) &= \lambda_1\left(s_u + \frac{\eta_u^{s,(n)}}{\sqrt{n}}, i_u + \frac{\eta_u^{i,(n)}}{\sqrt{n}}\right) - \lambda_1(s_u, i_u) \\ &= \int_0^1 \left\{ \partial_S \lambda_1\left(s_u + \alpha \frac{\eta_u^{s,(n)}}{\sqrt{n}}, i_u + \alpha \frac{\eta_u^{i,(n)}}{\sqrt{n}}\right) \frac{\eta_u^{s,(n)}}{\sqrt{n}} + \partial_I \lambda_1\left(s_u + \alpha \frac{\eta_u^{s,(n)}}{\sqrt{n}}, i_u + \alpha \frac{\eta_u^{i,(n)}}{\sqrt{n}}\right) \frac{\eta_u^{i,(n)}}{\sqrt{n}} \right\} d\alpha,\end{aligned}$$

and from the Lipschitz properties of  $\partial_S \lambda_1$  and  $\partial_I \lambda_1$  that  $|A(n, u)| \leq C(s_u, i_u)(|\eta_u^{s,(n)}|^2 + |\eta_u^{i,(n)}|^2)/\sqrt{n}$  and:

$$\mathbb{P}\left(\sup_{t \in [0, T]} \left| \int_0^t A(n, u) du \right| > \varepsilon\right) \leq \frac{CNE\left(\sup_{t \in [0, T \wedge \zeta_N^n]} |\eta_u^{s,(n)}|^2 + |\eta_u^{i,(n)}|^2\right)}{\varepsilon \sqrt{n}} + \mathbb{P}(\zeta_N^n \leq T), \quad (38)$$

which tends to zero as  $n \rightarrow \infty$ , by virtue of Lemma 5.1 and of (29). We deal with the term  $B(n, s)$  with similar computations and obtain that  $\int_0^t B(n, s) ds$  converges in probability to 0 uniformly in  $t \in [0, T]$ . As a consequence,  $\Psi(\eta^{(n)}, \phi, \cdot)$  converges uniformly in  $t \in [0, T]$  and in probability to the same limit as  $(\widetilde{M}_t^{s,(n)}, \widetilde{M}_t^{i,(n)}, \widetilde{M}_t^{r,(n)}(\phi))$ , which shows that the limiting values  $\eta$  satisfy (10).

In order to complete Theorem 3.4's proof, we establish that the strong uniqueness property holds for (10) (for given  $W$  and  $\eta_0$ ) and that its solution belongs to  $\mathcal{C}([0, T], \mathbb{R}^2 \times C^{-2,0})$ . These results rely on the use of Gronwall's lemma and on the fact that when the density dependence is 'frozen' ( $s^{(n)}, i^{(n)}$  or  $r^{(n)}(da)$  have been replaced by their deterministic limits) the constants appearing in the estimates do not depend on the localization in  $N$  any more.

Consequently, the adherence value of  $(\mathcal{L}(\eta^{(n)}))_{n \in \mathbb{N}^*}$  is unique and the sequence  $\eta^{(n)}$  converges in law in  $\mathbb{D}([0, T], \mathbb{R}^2 \times W^{-3,0})$  to the solution of SDE (10) in  $\mathcal{C}([0, T], \mathbb{R}^2 \times C^{-2,0})$ .

#### A4 Proof of Theorem 4.1

Using representation (13), the convergence (15) directly results from Theorem 3.2 combined with the assumed smoothness properties of  $\lambda_1, \lambda_3$ . Now, by virtue of the identifiability assumption, the limiting contrast  $K(\theta, \theta^*)$  equals to 0 in the sole case where  $\theta = \theta^*$  and (16) then follows from the regularity assumption **R1** in a standard fashion (see [16]).

#### A5 Proof of Theorem 4.2

Observe first that the map  $\theta \in \Theta \mapsto l_T^{(n)}(\theta)$  is twice differentiable, and denoting by  $\theta^* \in \Theta$  the true value of the parameter, for all  $\theta \in \Theta$  and  $T > 0$ , the score  $\nabla_\theta l_T^{(n)}(\theta)$  equals to

$$\begin{aligned}& \int_{t=0}^T \int_{u=0}^\infty \frac{\nabla_\theta \lambda_2(\theta)}{\lambda_2(\theta)} \mathbf{1}_{\{0 \leq u \leq n \lambda_2(\theta^*) i_{t-}^{(n)}\}} Q^I(dt, du) \\ & + \int_{t=0}^T \int_{u=0}^\infty \frac{\nabla_\theta \lambda_3(i_{t-}^{(n)}, \langle r_{t-}^{(n)}, \psi \rangle, \theta)}{\lambda_3(i_{t-}^{(n)}, \langle r_{t-}^{(n)}, \psi \rangle, \theta)} \mathbf{1}_{\{\lambda_2(\theta^*) n i_{t-}^{(n)} < u \leq \lambda_2 n i_{t-}^{(n)} + n \lambda_3(i_{t-}^{(n)}, \langle r_{t-}^{(n)}, \psi \rangle, \theta^*)\}} Q^I(dt, du) \\ & - n \int_{u=0}^T \{ \nabla_\theta \lambda_2(\theta^*) i_u^{(n)} + \nabla_\theta \lambda_3(i_u^{(n)}, \langle r_u^{(n)}, \psi \rangle, \theta) \} du\end{aligned}$$

Let us define  $I_{\theta^*}^{(n)} = -\mathcal{H}\theta l_T^{(n)}(\theta)$ . We have

$$I_{\theta^*}^{(n)} = -n \int_{u=0}^T \{ \mathcal{H}\theta \lambda_2(\theta) i_u^{(n)} + \mathcal{H}\theta \lambda_3(i_u^{(n)}, \langle r_u^{(n)}, \psi \rangle, \theta) \} du + \int_{t=0}^T \int_{u=0}^{\infty} \left[ \frac{\nabla_{\theta} \lambda_2(\theta) \cdot^t \nabla_{\theta} \lambda_2(\theta)}{\lambda_2(\theta)} i_{t-}^{(n)} \mathbf{1}_{\{0 \leq u \leq n \lambda_2 i_{t-}^{(n)}\}} \right. \\ \left. + \frac{\nabla_{\theta} \lambda_3(i_{t-}^{(n)}, \langle r_{t-}^{(n)}, \psi \rangle, \theta) \cdot^t \nabla_{\theta} \lambda_3(i_{t-}^{(n)}, \langle r_{t-}^{(n)}, \psi \rangle, \theta)}{\lambda_3(i_{t-}^{(n)}, \langle r_{t-}^{(n)}, \psi \rangle, \theta)} \mathbf{1}_{\{\lambda_2(\theta^*) n i_{t-}^{(n)} < u \leq \lambda_2(\theta^*) n i_{t-}^{(n)} + n \lambda_3(i_{t-}^{(n)}, \langle r_{t-}^{(n)}, \psi \rangle, \theta^*)\}} \right] Q^I(dt, du).$$

We have the following result.

LEMMA 5.3 *Under the assumptions of Theorem (4.2), we have:*

- (i)  $I_{\theta^*}^{(n)} \rightarrow \mathcal{I}_{\theta^*}$  in  $\mathbb{P}_{\theta^*}$ -probability, as  $n \rightarrow \infty$ ,
- (ii) for all  $T > 0$ , the sequence of processes  $(\{n^{-1/2} \nabla_{\theta} l_t^{(n)}(\theta)\}_{t \in [0, T]}, n \in \mathbb{N}^*)$ , converges in law in  $\mathbb{D}(\mathbb{R}_+, \mathbb{R})$  to the continuous gaussian martingale process with 0 as initial value and quadratic variation given by:

$$\int_0^t \left\{ \frac{\nabla_{\theta} \lambda_2(\theta^*) \cdot^t \nabla_{\theta} \lambda_2(\theta^*)}{\lambda_2(\theta^*)} i_u^* + \frac{\nabla_{\theta} \lambda_3(i_u^*, \langle r_u^*, \psi \rangle, \theta^*) \cdot^t \nabla_{\theta} \lambda_3(i_u^*, \langle r_u^*, \psi \rangle, \theta^*)}{\lambda_3(i_u^*, \langle r_u^*, \psi \rangle, \theta^*)} \right\} ds \quad (39)$$

*Proof* The first assertion follows from Theorem 3.2. The second one is a consequence of Theorem 3.4.  $\square$

The argument of the asymptotic normality results may be classically derived from the lemma above (see Chapter 4 in [23] for instance). Technical details are omitted.

## References

- [1] R.A. Adams, *Sobolev spaces*, Academic Press, 1975.
- [2] D. Aldous, *Stopping times and tightness*, Annals of Probability **6** (1978), 335–340.
- [3] R.M. Anderson and R.M.C. May, *Infectious diseases of humans: dynamics and control*, Oxford University Press, 1991.
- [4] H. Andersson and T. Britton, *Stochastic epidemic models and their statistical analysis*, Lecture Notes in Statistics, vol. 151, Springer, New York, 2000.
- [5] N.G. Becker, *Uses of the EM algorithm in the analysis of data on HIV/AIDS and other infectious diseases*, Statist. Meth. Med. Research **6** (1997), 164–171.
- [6] A. Beskos, O. Papaspiliopoulos, G. Roberts, and P. Fearnhead, *Exact and computationally efficient likelihood-based inference for discretely observed diffusion processes*, Journal of the Royal Statistical Society, Series B (Statistical Methodology) **68** (2006), 333–382.
- [7] F. Brauer and C. Castillo-Chávez, *Mathematical models in population biology and epidemiology*, Springer, 2001.
- [8] S. Busenberg and K.L. Cooke, *Vertically transmitted diseases*, Springer-Verlag, 1993.
- [9] P. Driessche D. Kirchner C. Castillo-Chavez, S. Blower and A.A. Yakubu, *Mathematical approaches for emerging and reemerging diseases*, New Trends in Stochastic Analysis (Charingworth, 1994), The IMA Volumes in Mathematics and its Applications, vol. 126.
- [10] H. de Arazoza, J. Joanes, R. Lounes, C. Legeai, S. Cléménçon, J. Perez, and B. Auvet, *The HIV/AIDS epidemic in Cuba: description and tentative explanation of its low prevalence*, BMC Disease (2007).
- [11] H. de Arazoza and R. Lounes, *A nonlinear model for sexually transmitted disease with contact-tracing*, J. Math. Appl. Med. Biol. **19** (2002), 221–234.
- [12] S.N. Ethier and T.G. Kurtz, *Markov processus, characterization and convergence*, John Wiley & Sons, New York, 1986.
- [13] N. Fournier and S. Méléard, *A microscopic probabilistic description of a locally regulated population and macroscopic approximations*, Ann. Appl. Probab. **14** (2004), no. 4, 1880–1919.
- [14] Y.H. Hsieh, H. de Arazoza, S.M. Lee, and C.W. Chen, *Estimating the number of Cubans infected sexually by human immunodeficiency virus using contact tracing data*, Int. J. Epidemiol. **31** (2002), no. 3, 679–83.
- [15] J.M. Hyman, J. Li, and E.A. Stanley, *Modeling the impact of random screening and contact tracing in reducing the spread of HIV*, Math. Biosc. (2003), no. 181, 17–54.
- [16] I.A. Ibragimov and R.Z. Hasminskii, *Statistical estimation. asymptotic theory*, Applications of Mathematics, vol. 16, Springer-Verlag, New York-Berlin, 1981, Translated from the Russian by Samuel Kotz.
- [17] E.L. Ionides, C. Bretó, and A.A. King, *Inference for nonlinear dynamical systems*, Proceedings of the National Academy of Sciences of the USA **103** (2006), 18438–18443.
- [18] Isham, *Stochastic models for epidemics with special reference to AIDS*, The Annals of Applied Probability **3** (1993), no. 1, 1–27.
- [19] J. Jacod and A.N. Shiryaev, *Limit theorems for stochastic processes*, Springer-Verlag, Berlin, 1987.
- [20] A. Joffe and M. Métivier, *Weak convergence of sequences of semimartingales with applications to multitype branching processes*, Advances in Applied Probability **18** (1986), 20–65.
- [21] W.O. Kermack and A.G. McKendrick, *Contributions to the mathematical theory of epidemics*, Bulletin of Mathematical Biology, Springer New York **53** (1991), no. 1-2, 33–55.
- [22] Y. Kuang, *Delay differential equations with applications to population dynamics*, Academic Press, San Diego, CA, 1994.
- [23] Y.N. Linkov, *Asymptotic statistical methods for stochastic processes*, American mathematical society ed., Translations of Mathematical Monographs, vol. 196, Providence, 2001, Translated from the 1993 Russian original by Victor Kotov.

- [24] S. Méléard, *Convergence of the fluctuations for interacting diffusions with jumps associated with Boltzmann equations*, Stochastics and Stochastics Reports **63** (1998), 195–225.
- [25] S. Méléard and S. Roelly, *Sur les convergences étroite ou vague de processus à valeurs mesures*, C.R.Acad.Sci.Paris, Serie I **317** (1993), 785–788.
- [26] M. Métivier, *Convergence faible et principe d'invariance pour des martingales à valeurs dans des espaces de Sobolev*, Annales de l'IHP **20** (1984), no. 4, 329–348.
- [27] S.P. Meyn and R.L. Tweedie, *Stability of markovian processes III: Foster-Lyapunov criteria for continuous-time processes*, Advances in Applied Probability **25** (1993), no. 3, 518–548.
- [28] J. Müller, M. Kretzschmar, and K. Dietz, *Contact tracing in stochastic and deterministic epidemic models*, Math. Biosc. (2000), no. 164, 39–64.
- [29] C.J. Mode and C.K. Sleeman, *Stochastic processes in epidemiology*, World Scientific, 2000.
- [30] S.T. Rachev, *Probability metrics and the stability of stochastic models*, John Wiley & Sons, 1991.
- [31] B. Rapatski, P. Klepac, S. Dueck, M. Liu, and L.I. Weiss, *Mathematical epidemiology of HIV/AIDS in Cuba during the period 1986–2000*, Math. Bios. Engin. **3** (2006), no. 3, 545–556.
- [32] G. Roberts and O. Stramer, *On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm*, Biometrika **88** (2001), 603–621.
- [33] S. Roelly, *A criterion of convergence of measure-valued processes: Application to measure branching processes*, Stochastics **17** (1986), 43–65.
- [34] V.C. Tran, *Modèles particuliers stochastiques pour des problèmes d'évolution adaptative et pour l'approximation de solutions statistiques*, Ph.D. thesis, Université Paris X - Nanterre, <http://tel.archives-ouvertes.fr/tel-00125100>.
- [35] ———, *Large population limit and time behaviour of a stochastic particle model describing an age-structured population*, ESAIM: P&S (2008), to appear, DOI:10.1051/ps:2007052.
- [36] P. van den Driessche, *Some epidemiological models with delays*, Differential Equations and Applications to Industry (Claremont, CA, 1994) (River Edge, NJ), World Sci. Publishing, 1996, pp. 507–520.
- [37] ———, *Time delay in epidemic models*, Mathematical Approaches for Emerging and Reemerging Infectious Diseases: An Introduction (P. van den Driessche D. Kirschner C. Castillo-Chavez, S. Blower and A.A. Yakubu, eds.), vol. 125, IMA Volumes in Mathematics and its Applications, 2002, pp. 119–128.
- [38] E. van Doorn, *Quasi-stationary distributions and convergence to quasi-stationarity of birth-death processes*, Advances in Applied Probability **23** (1991), 683–700.
- [39] G.F. Webb, *Theory of nonlinear age-dependent population dynamics*, Monographs and Textbooks in Pure and Applied mathematics, vol. 89, Marcel Dekker, inc., New York - Basel, 1985.
- [40] World Health Organization, *Web site and related documents*, <http://www.who.int/countries/cub/en/>, [http://data.unaids.org/pub/EpiReport/2006/20061121\\_EPI\\_FS\\_C\\_en.pdf](http://data.unaids.org/pub/EpiReport/2006/20061121_EPI_FS_C_en.pdf).