Research Article

Stochastic Methodology for the Study of an Epidemic Decay Phase, Based on a Branching Model

Sophie Pénisson[1] and Christine Jacob[2]

[1] *Université Paris-Est, LAMA (UMR 8050), UPEMLV, UPEC, CNRS, 94010 Créteil, France*
[2] *INRA, MIA (UR 341), 78352 Jouy-en-Josas, France*

Correspondence should be addressed to Sophie Pénisson, sophie.penisson@u-pec.fr

We present a stochastic methodology to study the decay phase of an epidemic. It is based on a general stochastic epidemic process with memory, suitable to model the spread in a large open population with births of any rare transmissible disease with a random incubation period and a Reed-Frost type infection. This model, which belongs to the class of multitype branching processes in discrete time, enables us to predict the incidences of cases and to derive the probability distributions of the extinction time and of the future epidemic size. We also study the epidemic evolution in the worst-case scenario of a very late extinction time, making use of the Q-process. We provide in addition an estimator of the key parameter of the epidemic model quantifying the infection and finally illustrate this methodology with the study of the Bovine Spongiform Encephalopathy epidemic in Great Britain after the 1988 feed ban law.

1. Introduction

Outbreaks of infectious diseases of animals or humans are subject, when possible, to control measures aiming at curbing their spread. Effective measures should force the epidemic to enter its decay phase and to reach extinction. The decay phase can then be simply detected by a decrease of the number of cases, when this decrease is obvious. However this is not always the case, and this rough qualitative information might not be sufficient to evaluate accurately the effectiveness of the proposed measures to reduce the final size and duration of the outbreak. The goal of this paper is to present a stochastic methodology in discrete time to study more accurately the decay phase of an epidemic. Our framework is the spread, in a large open population, of a rare transmissible disease such that the infection process may be

assumed to follow a Reed-Frost type model, with a probability for a susceptible to become infected by a given dose of pathogens inversely proportional to the total population size. Moreover the latent period (during which an individual is infected but not yet infectious) may be random and long compared to the generation time. Questions about the decay phase include the following: which quantitative criteria can ensure that the disease has entered an extinction phase? What is the probability distribution of the epidemic extinction time, of the epidemic final size, and of the incidence of infected individuals? Finally, what would be the evolution of the epidemic in the event of a very late extinction of the disease?

From a practical point of view, it is generally impossible to observe all infections. Susceptible and infected but not yet infectious individuals are most often not distinguishable, being both apparently healthy. This leads to the fact that the only available observations correspond to the incidence of individuals with clinical symptoms. One way to deal with this lack of information was proposed in [1] by Panaretos, who used a model taking into account two types of infected individuals, the observed and the unobserved. In order to answer the previous questions, we choose here a different approach, considering a stochastic model depending on the sole incidences $\{X_n\}_n$ of infectives at each time. We assume that an infective can transmit the disease during one given time unit at most. Therefore the incidence of infectives corresponds to the incidence of cases. The process then describes in a recursive way how one single infective can indirectly generate new infectives (so-called "secondary cases") k time units later, where $1 \leqslant k \leqslant d$. We assume that this number of secondary cases follows a Poisson probability distribution with parameter $\Psi_k > 0$. The recursive formula defining $\{X_n\}_n$ is then the following:

$$X_n = \sum_{k=1}^{d} \sum_{i=1}^{X_{n-k}} Y_{n-k,n,i}, \tag{1.1}$$

where the variable $Y_{n-k,n,i}$ is the incidence of secondary cases produced at time n with a delay k (latent period) by individual i infectious at time $n - k$. The $\{Y_{n-k,n,i}\}_{i,k}$ are assumed independent given $\mathcal{F}_{n-1} := \sigma(\{X_{n-k}\}_{k \geqslant 1})$, and the $\{Y_{n-k,n,i}\}_i$ are assumed i.i.d. (identically and independently distributed) given $\mathcal{F}_{n-1}$, with a common Poisson distribution with parameter Ψ_k. This model is therefore time homogeneous and is in this sense less general than the one introduced in [2], which describes the spread of infectious animal diseases in a varying environment. However since we focus on the extinction phase only, the assumption of a constant environment with no new control measure is well founded and enables us to describe more accurately the decay phase. This process is the generalization of the well-known single-type BGW (Bienaymé-Galton-Watson) branching process, which is the limit, as the total population size tends to infinity, of the process describing the spread in a closed population of an infectious disease with a negligible latent period and a probability to become infected following a Reed-Frost model (see, e.g., [3, 4] and citations therein).

The core of the paper lies in Section 2, where the whole methodology is presented. We first formulate the epidemic model $\{X_n\}_n$ as a multitype branching process with Poissonian transitions, the types representing the memory of the process. This formulation provides useful analytical results such as the extinction criteria, and the distributions of the extinction time and of the epidemic size (Sections 2.1–2.3). Then, in order to investigate the worst-case scenario of an extreme late extinction of the epidemic, we introduce in Section 2.4 the Q-process $\{X_n^*\}_n$, obtained by conditioning $\{X_n\}_n$ on a very late extinction. Using this process, we focus the study on the early behavior of the decay phase in the worst-case scenario, rather

than on its long range behavior, which would have little meaning in our setting. Motivated by practical applications to real epidemics, for which we want to predict the processes $\{X_n\}_n$ and $\{X_n^*\}_n$, as well as the derived distributions above, we need to know the values of the parameters $\{\Psi_k\}_k$. We may write $\Psi_k = \sum_{a=1}^{a_m-k} \theta_a P_{\text{inc},a}(k) P_{\text{age}}(a+k)$, where θ_a is the mean number of individuals infected at age a by an infective by direct or indirect transmission, a_m is the largest survival age, $P_{\text{inc},a}(k)$ is the probability for the individual aged a at infection to have a latent period equal to k (given his survival), and $P_{\text{age}}(a+k)$ is the probability to be aged $a + k$ at the end of the latent period. Parameters $\{\theta_a\}_a$ are the key quantities for the spread of the disease and can be subject to changes due to control measures during the epidemic. We assume here that $\theta_a = \theta_0 + p_a$, where p_a is constant over time, while θ_0 may change with control measures. A typical example is when θ_0 is the mean number of individuals infected by an infective by horizontal route at age a, assumed independent of $a \geqslant 2$, and p_1 represents the maternal transmission probability p_{mat}. In this case

$$\Psi_k = \theta_0 \sum_{a=k+1}^{a_m} P_{\text{inc},a-k}(k) P_{\text{age}}(a) + p_{\text{mat}} P_{\text{inc},1}(k) P_{\text{age}}(k+1). \tag{1.2}$$

So we assume here that, except for θ_0, the other parameters of the $\{\Psi_k\}_k$ are constant over time and are known (generally estimated) from previous experiences or from the study of the whole epidemic evolution, in particular its growth phase. We moreover assume that each Ψ_k depends affinely on θ_0 (see Section 3.1). In Sections 2.5 and 2.6, we provide optimal WCLSE (Weighted Conditional Least Squares Estimators) of θ_0 in the decay phase, in the frame of $\{X_n\}_n$ as well as in the frame of the associated conditioned process $\{X_n^*\}_n$, and prove the strong consistency and the asymptotic normality of these estimators.

The final Section 3 is devoted to the application of this method to real epidemics. We first present in Section 3.1 some general conditions under which the spread of a SEIR disease (susceptible, exposed (latent), infectious, removed) can be approximated by our epidemic process defined by (1.1) and give an explicit derivation of the parameters $\{\Psi_k\}_k$. We then illustrate the methodology in Section 3.2 with the decay phase of the BSE (Bovine Spongiform Encephalopathy) epidemic in Great Britain. According to the available data [5], the epidemic is obviously fading out. We assume that the $\{\Psi_k\}_k$ satisfy (1.2). Then thanks to the stochastic tools developed here, we provide in addition to this rough information, short- and long-term predictions about the future spread of the disease as well as an estimation of a potentially remaining horizontal infection route after the 1988 feed ban law.

2. Methodology for the Study of an Epidemic Decay Phase

In this section we present a general methodology to study the decay phase of a SEIR disease in a large population, modeled by the process (1.1) defined in Section 1. Our main goal is to provide analytical tools to evaluate the efficiency of the last control measures taken prior to the considered time period. Most of our results are derived from the fact that this epidemic model can be seen as a multitype branching process. Indeed, $\{X_n\}_n$ defined by (1.1) is a Markovian process of order d. Consequently, the d-dimensional process $\{\mathbf{X}_n\}_n$ defined by

$$\mathbf{X}_n := (X_n, X_{n-1}, \dots, X_{n-d+1}) \tag{2.1}$$

is Markovian of order 1, and it stems directly from (1.1) that $\{\mathbf{X}_n\}_n$ is a multitype Bienaymé-Galton-Watson (BGW) process with d types (see, e.g., [6]). Note that the d types in this branching process do not correspond to any attribute of the individuals in the population, which is usually the case in mathematical biology (see, e.g., [7]), but simply correspond to the memory of the process $\{X_n\}_n$. The information provided by the d-dimensional Markovian process $\{\mathbf{X}_n\}_n$ is therefore the same as the one given by the 1-dimensional d-Markovian process $\{X_n\}_n$, but the multitype branching process setting gives us powerful mathematical tools and results stemming from the branching processes theory [6]. The first basic tool is the generating function of the offspring distribution of $\{\mathbf{X}_n\}_n$, $\mathbf{f} := (f_1, \ldots, f_d)$, defined on $[0,1]^d$ by $f_i(\mathbf{r}) := \mathbb{E}[\mathbf{r}^{\mathbf{X}_1} \mid \mathbf{X}_0 = \mathbf{e}_i]$, where $\mathbf{e}_i := (0, \ldots, 1, \ldots, 0)$ denotes the ith basis vector of $\mathbb{N}^d$ and $\mathbf{u}^{\mathbf{v}} := \prod_{i=1}^d u_i^{v_i}$ for $\mathbf{u}, \mathbf{v} \in \mathbb{N}^d$. For all $\mathbf{r} \in [0,1]^d$, we have here

$$f_i(\mathbf{r}) := e^{-(1-r_1)\Psi_i} r_{i+1}, \quad i = 1 \cdots d-1,$$

$$f_d(\mathbf{r}) := e^{-(1-r_1)\Psi_d}.$$

(2.2)

The second basic tool is the mean matrix $\mathbf{M}$ defined by $\mathbb{E}(\mathbf{X}_n \mid \mathcal{F}_{n-1}) = \mathbf{X}_{n-1}\mathbf{M}$, which is here

$$\mathbf{M} = \begin{pmatrix} \Psi_1 & 1 & 0 & \cdots & 0 \\ \Psi_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ \Psi_{d-1} & 0 & \cdots & \cdots & 1 \\ \Psi_d & 0 & \cdots & \cdots & 0 \end{pmatrix}.$$

(2.3)

Let us notice that, since $\Psi_k > 0$ for each $k = 1 \cdots d$, then $\{\mathbf{X}_n\}_n$ is nonsingular, positive regular (see [6]) and satisfies the $X \log X$ condition,

$$\mathbb{E}[\|\mathbf{X}_1\| \ln \|\mathbf{X}_1\| \mid \mathbf{X}_0 = \mathbf{e}_i] < \infty, \quad i = 1, \ldots, d,$$

(2.4)

where $\|\cdot\|$ denotes the sup norm in $\mathbb{R}^d$.

2.1. Extinction of the Epidemic

2.1.1. Almost Sure Extinction

Since the single-type process $\{X_n\}_n$ has a memory of size d, it becomes extinct when it is null at d successive times, or equivalently as soon as the d-dimensional process $\{\mathbf{X}_n\}_n$ reaches the d-dimensional null vector $\mathbf{0}$. According to the theory of multitype positive regular and nonsingular BGW processes [6], the extinction of the process $\{\mathbf{X}_n\}_n$ occurs almost surely (a.s.), if and only if $\rho \leqslant 1$, where ρ is the dominant eigenvalue (also called the Perron's root) of the mean matrix $\mathbf{M}$. Thus ρ is solution of $\sum_{k=1}^d \Psi_k \rho^{-k} = 1$. In general for $d > 1$, ρ has no explicit expression. However, $\sum_{k=1}^d \Psi_k \rho^{-k} = 1$ leads directly to the following explicit threshold criteria.

Proposition 2.1. *The epidemic becomes extinct almost surely if and only if $R_0 \leqslant 1$, where $R_0 := \sum_{k=1}^d \Psi_k$ is the total mean number of secondary cases generated by one infective in a SEIR disease. We call R_0 the basic reproduction number.*

Moreover, when $\rho \leqslant 1$, then $R_0 \leqslant \rho$ with equality if and only if either $\rho = 1$ or $d = 1$, and when $\rho > 1$, then $R_0 \geqslant \rho$ with equality if and only if $d = 1$.

Note that when $d > 1$, R_0 only provides information about the threshold level, whereas ρ provides an additional information about the speed of extinction of the process, as shown in the next two paragraphs.

2.1.2. Speed of Extinction

Thanks to well-known results in the literature about multitype branching processes and more particularly to the Perron-Frobenius theorem (see, e.g., [6]), we can deduce the expected incidence of infectives in the population at time n, for n large. Denoting by $\mathbf{u}$ and $\mathbf{v}$ the right and left eigenvectors of $\mathbf{M}$ associated to the Perron's root ρ; that is, $\mathbf{Mu}^T = \rho \mathbf{u}^T$ and $\mathbf{vM} = \rho \mathbf{v}$, with the normalization convention $\mathbf{u} \cdot \mathbf{1} = \mathbf{u} \cdot \mathbf{v} = 1$, where $\mathbf{u} \cdot \mathbf{v}$ stands for the usual scalar product in $\mathbb{R}^d$ and where the superscript T denotes the transposition, then $\mathbb{E}(\mathbf{X}_n \mid \mathbf{X}_0) = \mathbf{X}_0 \mathbf{M}^n \underset{n \to \infty}{\sim} \rho^n \mathbf{X}_0 \mathbf{u}^T \mathbf{v}$. The first coordinate in the latter formula becomes for the epidemic process $\rho^n \sum_{i=1}^d X_{-i+1} u_i v_1$. Computing explicitly $\mathbf{u}$ and $\mathbf{v}$, we obtain that for all $i = 1 \cdots d$,

$$u_i = \frac{\sum_{k=i}^d \Psi_k \rho^{-k+i}}{\sum_{j=1}^d \sum_{k=j}^d \Psi_k \rho^{-k+j}}, \qquad v_i = \rho^{-i} \frac{\sum_{j=1}^d \sum_{k=j}^d \Psi_k \rho^{-k+j}}{\sum_{j=1}^d \sum_{k=j}^d \Psi_k \rho^{-k}}, \tag{2.5}$$

which leads to the following asymptotic result:

$$\mathbb{E}(X_n \mid \mathbf{X}_0) \underset{n \to \infty}{\sim} \rho^n \sum_{i=1}^d X_{-i+1} \frac{\sum_{k=i}^d \Psi_k \rho^{-k+i-1}}{\sum_{j=1}^d \sum_{k=j}^d \Psi_k \rho^{-k}}. \tag{2.6}$$

Hence if $\rho < 1$, the mean number of infectives decreases exponentially at the rate ρ. In the following section, we provide a much finer result on the estimation of the disease extinction time in the population.

2.1.3. Extinction Time of the Epidemic

The extinction time distribution can be derived as a function of the offspring generating function. As usual in stochastic processes, this quantity is calculated conditionally on the initial value $\mathbf{X}_0 = (X_0, X_{-1}, \ldots, X_{-d+1})$, but for the sake of simplicity we do not let it appear in the notations. Note that since we are building tools for the prediction of the spread of the disease, the time origin 0 corresponds here to the time of the last available data (generally the current date). Let $T := \inf\{n \geqslant 1, \mathbf{X}_n = \mathbf{0}\}$ denote the extinction time of the process $\{\mathbf{X}_n\}_n$, and let $\mathbf{f}_n := \mathbf{f} \circ \mathbf{f}_{n-1}$ be the nth iterate of the generating function $\mathbf{f}$ given by (2.2). We denote $\mathbf{f}_n := (f_{n,1}, \ldots, f_{n,d})$. Then, by the branching property of the process ([6]), the probability of extinction of the epidemic before time n is immediately given by $\mathbb{P}(T \leqslant n) = \mathbb{P}(\mathbf{X}_n = \mathbf{0}) = \mathbf{f}_n(\mathbf{0})^{\mathbf{X}_0}$, that is to say,

$$\mathbb{P}(T \leqslant n) = \left(f_{n,1}(\mathbf{0})\right)^{X_0} \left(f_{n,2}(\mathbf{0})\right)^{X_{-1}} \cdots \left(f_{n,d}(\mathbf{0})\right)^{X_{-d+1}}. \tag{2.7}$$

It can be immediately deduced from convergence results for $\mathbf{f}_n(\mathbf{0})$ as $n \to \infty$ [8], that if $\rho = 1$, $\mathbb{P}(T \leqslant n) \sim 1 - (n\eta)^{-1}\mathbf{X}_0 \cdot \mathbf{u}$, while if $\rho < 1$, $\mathbb{P}(T \leqslant n) \sim 1 - \rho^n\gamma\mathbf{X}_0 \cdot \mathbf{u}$, for some constants $\eta, \gamma > 0$. As a consequence, the closer ρ is to unity, the longer the time to extinction will be in most realizations. More specifically, (2.7) enables the exact computation (resp., estimation) of $\mathbb{P}(T \leqslant n)$ for any n by the iterative computation of $\mathbf{f}_n$, $\mathbf{X}_0$ being given, when the parameters Ψ_k of (1.1) are known (resp., estimated). Moreover, since for $\rho \leqslant 1$ the epidemic becomes extinct in an a.s. finite time and $\mathbb{P}(T \leqslant n) \nearrow_{n\to\infty} 1$, then for any given probability $p \in [0, 1[$ there exists $n \in \mathbb{N}$ such that $\mathbb{P}(T \leqslant n) \geqslant p$. So in practice, for any $p \in [0, 1[$, (2.7) enables us to compute the p-quantile n_p^T of the extinction time,

$$n_p^T := \min\{n \geqslant 1 : \mathbb{P}(T \leqslant n) \geqslant p\}. \tag{2.8}$$

2.2. Total Size of the Epidemic

Under the assumption $\rho \leqslant 1$ and the independence of the $\{Y_{n-l,n,i}\}_{i,l,n}$ (we previously assumed the independence of the $\{Y_{n-l,n,i}\}_{i,l}$, for each n), we derive the distribution of the future total size $N := \sum_{n=1}^{T} X_n$ of the epidemic until its extinction, that is, the future total number of infectives until the extinction of the disease. It can be shown [9] that, given the initial value $\mathbf{X}_0$, the time origin being the same as in Section 2.1, the probability distribution of N is

$$N \overset{\mathfrak{D}}{=} \sum_{k=1}^{d} \sum_{i=1}^{X_{-k+1}} \left(\sum_{j=1}^{Y_{k,i}} N_{k,i,j} \right), \qquad Y_{k,i} := \sum_{l=k}^{d} Y_{-k+1,-k+1+l,i}, \tag{2.9}$$

where $\overset{\mathfrak{D}}{=}$ denotes the equality in distribution, an empty sum is by convention 0, the $\{Y_{k,i}\}_{k,i}$ are independent, and the $\{Y_{k,i}\}_i$ and the $\{N_{k,i,j}\}_{k,i,j}$ are i.i.d. with

$$Y_{k,i} \overset{\mathfrak{D}}{=} \mathrm{Poiss}\left(\sum_{l=k}^{d} \Psi_l \right), \qquad N_{k,i,j} \overset{\mathfrak{D}}{=} \mathrm{Borel} - \mathrm{Tanner}\left(\sum_{l=1}^{d} \Psi_l, 1 \right), \tag{2.10}$$

that is, for each $n \geqslant 1$, $\mathbb{P}(N_{k,i,j} = n) = e^{-n\sum_{l=1}^{d}\Psi_l}(n\sum_{l=1}^{d}\Psi_l)^{n-1}(n!)^{-1}$. Consequently, under the convention that an empty product is 1, the probability distribution of N is, for any $n \in \mathbb{N}$,

$$\mathbb{P}(N = n) = \sum_{\substack{\{0 \leqslant y_{k,i} \leqslant n, \{1 \leqslant n_{k,i,j} \leqslant n\}_j\}_{i,k}: \\ \sum_{k=1}^{d}\sum_{i=1}^{X_{-k+1}}\sum_{j=1}^{y_{k,i}} n_{k,i,j}=n}} \prod_{k=1}^{d} \prod_{i=1}^{X_{-k+1}} e^{-\sum_{l=k}^{d}\Psi_l} \frac{\left(\sum_{l=k}^{d}\Psi_l\right)^{y_{k,i}}}{y_{k,i}!}$$

$$\times \prod_{j=1}^{y_{k,i}} e^{-n_{k,i,j}\sum_{l=1}^{d}\Psi_l} \frac{\left(n_{k,i,j}\sum_{l=1}^{d}\Psi_l\right)^{n_{k,i,j}-1}}{n_{k,i,j}!}, \tag{2.11}$$

which may be calculated (resp., estimated), replacing the Ψ_k by their values (resp., estimations). In practice, for any $p \in [0, 1[$, (2.11) enables to compute the p-quantile n_p^N of the total epidemic size,

$$n_p^N := \min\{n \geqslant 1 : \ \mathbb{P}(N \leqslant n) \geqslant p\}. \tag{2.12}$$

We obtain moreover an explicit formula for the mean value and variance of the size of the epidemic,

$$\mathbb{E}(N) = \frac{\sum_{k=1}^{d} X_{-k+1} \sum_{l=k}^{d} \Psi_l}{1 - \sum_{l=1}^{d} \Psi_l}, \qquad \mathbb{V}\mathrm{ar}(N) = \frac{\sum_{k=1}^{d} X_{-k+1} \sum_{l=k}^{d} \Psi_l}{\left(1 - \sum_{l=1}^{d} \Psi_l\right)^3}. \tag{2.13}$$

2.3. Exposed Population

Depending on the disease, it might also be crucial to study and predict the evolution of the incidence of exposed individuals in the population, which is generally unobservable. We assume that this information is given by the process $\{Z_n\}_n$ defined by the conditional distribution $Z_n \mid X_n \overset{\mathcal{D}}{=} \mathrm{Poiss}(\Psi_0 X_n)$, where Ψ_0 is the mean number of individuals infected at time n by an infective of this time (see Section 3.1). This property enables on the one hand to reconstruct the whole past epidemic (i.e., the incidence of infectives as well as of exposed individuals) thanks to the observable data. On the other hand, it allows to simulate the evolution of the incidence of exposed individuals in the future, based on predictions of the evolution of the epidemic process $\{X_n\}_n$.

2.4. Worst-Case Scenario: Very Late Extinction of the Epidemic

Even in the case when the epidemic dies out almost surely ($\rho \leqslant 1$), and although one can provide the p-quantile n_p^T of the extinction time with the probability p as large as wanted (see (2.8)), the epidemic might become extinct after this time with a small but nonnull probability of order $1 - p$. This raises the following question: how would the incidences of infectious and exposed individuals evolve in the (unlikely) case of a very late extinction? In terms of risk analysis, this issue appears to be crucial to evaluate the risks associated with this worst-case scenario. The tools developed in the previous subsections allow to evaluate the probability of all possible outcomes. But since the worst ones, typically a very late extinction, have a negligible probability, these tools do not bring any information in these worst cases and in particular do not inform on the evolution at each time step of the spread of the disease (would it decrease extremely slowly, stay at a constant rate for a very long time, present several peaks in its evolution, etc.?). In order to study the propagation of the epidemic in the decay phase, assuming that extinction occurs very late, we are interested in the distribution of the process $\{X_n\}_n$ conditionally on the event that the epidemic has not become extinct at time k, where k is very large. We therefore consider for any $n_1, n_2, \ldots \in \mathbb{N}$ and any $\mathbf{i}_0, \mathbf{i}_1, \mathbf{i}_2, \ldots \in \mathbb{N}^d$ the conditioned probability $\mathbb{P}_{\mathbf{i}_0}(\mathbf{X}_{n_1} = \mathbf{i}_1, \ldots, \mathbf{X}_{n_r} = \mathbf{i}_r \mid \mathbf{X}_k \neq \mathbf{0})$, where the subscript $\mathbf{i}_0$ denotes the initial value. If k is finite this distribution cannot be easily handled due to its time

inhomogeneity. However, when $\rho \leqslant 1$, it is known ([10]) that this conditioned distribution converges, as $k \to \infty$, to the distribution of a d-dimensional Markov process $\{\mathbf{X}_n^*\}_n$:

$$\lim_{k \to \infty} \mathbb{P}_{\mathbf{i}_0}(\mathbf{X}_{n_1} = \mathbf{i}_1, \ldots, \mathbf{X}_{n_r} = \mathbf{i}_r \mid \mathbf{X}_k \neq \mathbf{0}) = \mathbb{P}_{\mathbf{i}_0}(\mathbf{X}_{n_1}^* = \mathbf{i}_1, \ldots, \mathbf{X}_{n_r}^* = \mathbf{i}_r). \tag{2.14}$$

We will further discuss in Proposition 2.5 the relevancy of approximating the conditioned probability for k fixed by the limiting object (2.14). The conditioned process $\{\mathbf{X}_n^*\}_n$ defined by (2.14) is known in the literature as the Q-process associated with $\{\mathbf{X}_n\}_n$, also described as the process conditioned on "not being extinct in the distant future." It has the following transition probability [10]: for every $n \geqslant 1$, $\mathbf{i}, \mathbf{j} \in \mathbb{N}^d$, $\mathbf{i} \neq \mathbf{0}$,

$$\mathbb{P}(\mathbf{X}_n^* = \mathbf{j} \mid \mathbf{X}_{n-1}^* = \mathbf{i}) = \frac{1}{\rho} \frac{\mathbf{j} \cdot \mathbf{u}}{\mathbf{i} \cdot \mathbf{u}} \mathbb{P}(\mathbf{X}_n = \mathbf{j} \mid \mathbf{X}_{n-1} = \mathbf{i}), \tag{2.15}$$

where $\mathbf{u}$ is the normalized right eigenvector of $\mathbf{M}$ associated to the Perron's root ρ as introduced in Section 2.1, and computed explicitly in (2.5). In the same way as for the process $\{\mathbf{X}_n\}_n$, we define the 1-dimensional process $X_n^* := X_{n,1}^*$. By construction we then have $X_{n,i}^* = X_{n-i+1}^*$, for each n and each $i = 1 \cdots d$.

Proposition 2.2. *The stochastic process $\{X_n^*\}_n$ is, conditionally on its past, distributed as the sum of two independent Poisson and Bernoulli random variables:*

$$X_n^* \mid \mathbf{X}_{n-1}^* \overset{\mathfrak{D}}{=} \mathrm{Poiss}(\mathbf{X}_{n-1}^* \cdot \mathbf{\Psi}) * \mathcal{B}(p(\mathbf{X}_{n-1}^*)), \tag{2.16}$$

where $\mathbf{\Psi} := (\Psi_1, \ldots, \Psi_d)$, $*$ is the convolution product symbol, and

$$p(\mathbf{X}_{n-1}^*) := \frac{u_1 \mathbf{X}_{n-1}^* \cdot \mathbf{\Psi}}{u_1 \mathbf{X}_{n-1}^* \cdot \mathbf{\Psi} + \sum_{k=2}^{d} X_{n-k+1}^* u_k}. \tag{2.17}$$

Proof. Applying (1.1) and (2.15), we obtain that for all $j \in \mathbb{N}$,

$$
\begin{aligned}
\mathbb{P}&(X_n^* = j \mid \mathbf{X}_{n-1}^*) \\
&= \mathbb{P}\left(\mathbf{X}_n^* = \left(j, X_{n-1}^*, \ldots, X_{n-(d-1)}^*\right) \mid \mathbf{X}_{n-1}^*\right) \\
&= \frac{ju_1 + \sum_{k=2}^{d} X_{n-k+1}^* u_k}{\rho \mathbf{X}_{n-1}^* \cdot \mathbf{u}} \mathbb{P}\left(\mathbf{X}_n = \left(j, X_{n-1}^*, \ldots, X_{n-(d-1)}^*\right) \mid \mathbf{X}_{n-1} = \mathbf{X}_{n-1}^*\right) \\
&= \frac{ju_1 + \sum_{k=2}^{d} X_{n-k+1}^* u_k}{\rho \mathbf{X}_{n-1}^* \cdot \mathbf{u}} \frac{(\mathbf{X}_{n-1}^* \cdot \mathbf{\Psi})^j}{j!} e^{-\mathbf{X}_{n-1}^* \cdot \mathbf{\Psi}} \\
&= \frac{u_1 \mathbf{X}_{n-1}^* \cdot \mathbf{\Psi}}{\rho \mathbf{X}_{n-1}^* \cdot \mathbf{u}} \frac{(\mathbf{X}_{n-1}^* \cdot \mathbf{\Psi})^{j-1}}{(j-1)!} e^{-\mathbf{X}_{n-1}^* \cdot \mathbf{\Psi}} + \frac{\sum_{k=2}^{d} X_{n-k+1}^* u_k}{\rho \mathbf{X}_{n-1}^* \cdot \mathbf{u}} \frac{(\mathbf{X}_{n-1}^* \cdot \mathbf{\Psi})^j}{j!} e^{-\mathbf{X}_{n-1}^* \cdot \mathbf{\Psi}}.
\end{aligned}
\tag{2.18}
$$

The equality $\mathbf{M}\mathbf{u}^T = \rho\mathbf{u}^T$ implies that for all $k = 1\cdots d - 1$, $\rho u_k = \Psi_k u_1 + u_{k+1}$, and that $\rho u_d = \Psi_d u_1$. Consequently, $\rho\mathbf{X}^*_{n-1} \cdot \mathbf{u} = u_1\mathbf{X}^*_{n-1} \cdot \mathbf{\Psi} + \sum_{k=2}^{d} X^*_{n-k+1} u_k$, and thus

$$
\begin{aligned}
\mathbb{P}&(X^*_n = j \mid \mathbf{X}^*_{n-1}) \\
&= \frac{u_1\mathbf{X}^*_{n-1} \cdot \mathbf{\Psi}}{u_1\mathbf{X}^*_{n-1} \cdot \mathbf{\Psi} + \sum_{k=2}^{d} X^*_{n-k+1} u_k} \frac{(\mathbf{X}^*_{n-1} \cdot \mathbf{\Psi})^{j-1}}{(j-1)!} e^{-\mathbf{X}^*_{n-1} \cdot \mathbf{\Psi}} \\
&\quad + \left(1 - \frac{u_1\mathbf{X}^*_{n-1} \cdot \mathbf{\Psi}}{u_1\mathbf{X}^*_{n-1} \cdot \mathbf{\Psi} + \sum_{k=2}^{d} X^*_{n-k+1} u_k}\right) \frac{(\mathbf{X}^*_{n-1} \cdot \mathbf{\Psi})^{j}}{j!} e^{-\mathbf{X}^*_{n-1} \cdot \mathbf{\Psi}} \\
&= \left[\mathrm{Poiss}(\mathbf{X}^*_{n-1} \cdot \mathbf{\Psi}) * \mathcal{B}\left(\frac{u_1\mathbf{X}^*_{n-1} \cdot \mathbf{\Psi}}{u_1\mathbf{X}^*_{n-1} \cdot \mathbf{\Psi} + \sum_{k=2}^{d} X^*_{n-k+1} u_k}\right)\right](j).
\end{aligned}
\tag{2.19}
$$

$\square$

Remark 2.3. Note that if one compares (2.16) with the transition probability of the unconditioned process $X_n \mid \mathbf{X}_{n-1} \stackrel{\mathcal{D}}{=} \mathrm{Poiss}(\mathbf{X}_{n-1} \cdot \mathbf{\Psi})$, it appears that $\{X^*_n\}_n$ behaves at each time step like $\{X_n\}_n$, according to a Poisson distribution, except that it has the possibility at each time step to add one unit or not, according to a Bernoulli random variable. Moreover, if $X^*_{n-1} = \cdots = X^*_{n-(d-1)} = 0$, then according to (2.17), $p(\mathbf{X}^*_{n-1}) = 1$, which implies that at time n, the probability to add one unit is equal to one, thus preventing the extinction of the process.

Proposition 2.4. *The process $\{\mathbf{X}^*_n\}_n$ admits a stationary probability measure π with finite first- and second-order moments.*

Proof. Since the multitype branching process $\{\mathbf{X}_n\}_n$ satisfies property (2.4), it is known [10] that the Q-process $\{\mathbf{X}^*_n\}_n$ is positive recurrent with a stationary probability measure π given by

$$
\pi(\mathbf{i}) := \frac{\mathbf{i} \cdot \mathbf{u}\ \nu(\mathbf{i})}{\sum_{\mathbf{k}\in\mathbb{N}^d} \mathbf{k} \cdot \mathbf{u}\ \nu(\mathbf{k})}, \quad \mathbf{i} \in \mathbb{N}^d,
\tag{2.20}
$$

where ν is the Yaglom distribution of the process $\{\mathbf{X}_n\}_n$, uniquely defined by the following property: for all $\mathbf{i}, \mathbf{j} \in \mathbb{N}^d \setminus \{\mathbf{0}\}$, $\lim_{n\to\infty}\mathbb{P}(\mathbf{X}_n = \mathbf{i} \mid \mathbf{X}_0 = \mathbf{j}, \mathbf{X}_n \neq \mathbf{0}) = \nu(\mathbf{i})$. In the literature, this stationary measure for the conditioned process $\{\mathbf{X}^*_n\}_n$ is also referred to as the doubly limiting conditional probability. Moreover, by Proposition 2.2,

$$
\mathbb{E}(X^*_n) = \mathbb{E}\big[\mathbb{E}(X^*_n \mid \mathbf{X}^*_{n-1})\big] = \mathbb{E}\big[\mathbf{X}^*_{n-1} \cdot \mathbf{\Psi} + p(\mathbf{X}^*_{n-1})\big] \leqslant \sum_{k=1}^{d} \mathbb{E}(X^*_{n-k})\Psi_k + 1,
\tag{2.21}
$$

which implies that $\overline{\lim}_{n\to\infty}\mathbb{E}(X^*_n) \leqslant (1 - \sum_{k=1}^{d} \Psi_k)^{-1} < \infty$. We consequently obtain by means of Fatou's lemma that, for every $i = 1\cdots d$,

$$
\sum_{\mathbf{j}\in\mathbb{N}^d} j_i\pi(\mathbf{j}) = \mathbb{E}\left(\lim_{n\to\infty} X^*_{n,i}\right) = \mathbb{E}\left(\lim_{n\to\infty} X^*_{n-i+1}\right) = \mathbb{E}\left(\lim_{n\to\infty} X^*_n\right) \leqslant \underline{\lim}_{n\to\infty} \mathbb{E}(X^*_n) < \infty.
\tag{2.22}
$$

We similarly prove that π has finite second-order moments by writing

$$\mathbb{V}\mathrm{ar}\left(X_n^* \mid X_{n-1}^*\right) = \mathbf{X}_{n-1}^* \cdot \mathbf{\Psi} + \frac{u_1 \mathbf{X}_{n-1}^* \cdot \mathbf{\Psi} \sum_{k=2}^d X_{n-k+1}^* u_k}{\left(u_1 \mathbf{X}_{n-1}^* \cdot \mathbf{\Psi} + \sum_{k=2}^d X_{n-k+1}^* u_k\right)^2} \leqslant \mathbf{X}_{n-1}^* \cdot \mathbf{\Psi} + \frac{1}{4}. \qquad (2.23)$$

$\square$

Let us discuss the relevancy of approximating the epidemic process $\{\mathbf{X}_n\}_n$ conditioned on nonextinction at some finite time k, for k large, by the Q-process $\{\mathbf{X}_n^*\}_n$ obtained by letting $k \to \infty$. When considering the case of late extinction, one works under an hypothetical assumption based on the unknown future, hence in practice one does not focus on a specific value k for the survival of the disease in the population. We therefore might consider that k is chosen large enough such that the approximation of the process $\{\mathbf{X}_n\}_n$ conditioned on the event $\{\mathbf{X}_k \neq \mathbf{0}\}$ by the process $\{\mathbf{X}_n^*\}_n$ is valid. Of course, the order of magnitude of such k will depend on the rate of convergence of the conditioned process to $\{\mathbf{X}_n^*\}_n$.

Proposition 2.5. *Let $n_1 \leqslant \cdots \leqslant n_r \leqslant k$ and $\mathbf{i}_0, \ldots, \mathbf{i}_r \in \mathbb{N}^d \setminus \{\mathbf{0}\}$. Then the difference $|\mathbb{P}_{\mathbf{i}_0}(\mathbf{X}_{n_1} = \mathbf{i}_1, \ldots, \mathbf{X}_{n_r} = \mathbf{i}_r \mid \mathbf{X}_k \neq \mathbf{0}) - \mathbb{P}_{\mathbf{i}_0}(\mathbf{X}_{n_1}^* = \mathbf{i}_1, \ldots, \mathbf{X}_{n_r}^* = \mathbf{i}_r)|$ decreases, as $k \to \infty$, with $\max\{k^{s-1}(|\lambda|\rho^{-1})^{k/2}, \rho^{k/2}\}$, where λ is an eigenvalue of $\mathbf{M}$ such that $\rho > |\lambda| \geqslant |\lambda_3| \geqslant |\lambda_4| \geqslant \cdots$, with the λ_i being the other eigenvalues of $\mathbf{M}$. In case $|\lambda| = |\lambda_3|$ we stipulate that the multiplicity s of λ is at least as great as the multiplicity of λ_3.*

Proof. Thanks to (2.15) and to the Markov property together with the fact that $\mathbb{P}_{\mathbf{i}}(\mathbf{X}_n = \mathbf{0}) = \mathbf{f}_n(\mathbf{0})^{\mathbf{i}}$, we have

$$\left|\mathbb{P}_{\mathbf{i}_0}(\mathbf{X}_{n_1} = \mathbf{i}_1, \ldots, \mathbf{X}_{n_r} = \mathbf{i}_r \mid \mathbf{X}_k \neq \mathbf{0}) - \mathbb{P}_{\mathbf{i}_0}\left(\mathbf{X}_{n_1}^* = \mathbf{i}_1, \ldots, \mathbf{X}_{n_r}^* = \mathbf{i}_r\right)\right|$$

$$= \left|\frac{1 - \mathbf{f}_{k-n_r}(\mathbf{0})^{\mathbf{i}_r}}{1 - \mathbf{f}_k(\mathbf{0})^{\mathbf{i}_0}} - \frac{1}{\rho^{n_r}} \frac{\mathbf{i}_r \cdot \mathbf{u}}{\mathbf{i}_0 \cdot \mathbf{u}}\right| \mathbb{P}_{\mathbf{i}_0}(\mathbf{X}_{n_1} = \mathbf{i}_1, \ldots, \mathbf{X}_{n_r} = \mathbf{i}_r). \qquad (2.24)$$

The right term of (2.24) is known to converge to 0, as $k \to \infty$, thanks to the property that $\lim_k \mathbf{a}_k = \gamma \mathbf{u}$, for some $\gamma > 0$, where $\mathbf{a}_k := \rho^{-k}(\mathbf{1} - \mathbf{f}_k(\mathbf{0}))$ (see [8]). This stems from two convergences, namely, $\lim_k b_k = \gamma$, where $b_k := \rho^{-k}(\mathbf{v} \cdot (\mathbf{1} - \mathbf{f}_k(\mathbf{0})))$, and $\lim_k \mathbf{a}_k b_k^{-1} = \mathbf{u}$. Let us write $\mathbf{a}_k = \gamma \mathbf{u} + \varepsilon_k$, where $\lim_k \varepsilon_k = \mathbf{0}$. Since $b_k = \mathbf{v} \cdot \mathbf{a}_k$ and $\mathbf{u} \cdot \mathbf{v} = 1$, it comes

$$\mathbf{a}_k b_k^{-1} = \frac{\gamma \mathbf{u} + \varepsilon_k}{\gamma + \mathbf{v} \cdot \varepsilon_k} \sim_{k \to \infty} \mathbf{u} + \gamma^{-1}[\varepsilon_k - (\mathbf{v} \cdot \varepsilon_k)\mathbf{u}]. \qquad (2.25)$$

It thus appears that the rate of convergence of $\mathbf{a}_k$ to $\gamma \mathbf{u}$ is of the same order of magnitude as the one of $\mathbf{a}_k b_k^{-1}$ to $\mathbf{u}$. Let us determine this rate in an accurate way. We use the following inequality produced by Joffe and Spitzer in [8]: for each $k \geqslant n \geqslant 1$,

$$\left\|\mathbf{a}_k b_k^{-1} - \mathbf{u}\right\| = \left\|\frac{\mathbf{1} - \mathbf{f}_k(\mathbf{0})}{\mathbf{v} \cdot (\mathbf{1} - \mathbf{f}_k(\mathbf{0}))} - \mathbf{u}\right\| \leqslant \frac{2\delta_n + \sum_{j=k-n+1}^k \alpha_j}{1 - \delta_n - \sum_{j=k-n+1}^k \alpha_j}, \qquad (2.26)$$

where we are going to replace δ_n and α_n by some explicit formulae function of ρ and n. For this purpose, we use a detailed asymptotic behavior of $\mathbf{M}^k$, as $k \to \infty$, presented for instance in [11]; we have $\mathbf{M}^k = \rho^k \mathbf{R} + O(k^{s-1} |\lambda|^k)$, where $\mathbf{R} = \mathbf{u}^T \mathbf{v}$. For the sake of clarity the symbol $O(\cdot)$ will denote either a scalar or a matrix with all the entries satisfying the associated property. This implies the existence of some constant $a > 0$ such that, for all $k \in \mathbb{N}$, $(1 - \delta_k)\mathbf{R} \leqslant \rho^{-k}\mathbf{M}^k \leqslant (1 + \delta_k)\mathbf{R}$, where $\delta_k := ak^{s-1}(\rho^{-1}|\lambda|)^k$. Moreover, following [8], let us write, for all $\mathbf{r} \in [0,1]^d$, $\mathbf{1} - \mathbf{f}(\mathbf{r}) = (\mathbf{M} - \mathbf{E}(\mathbf{r}))(\mathbf{1} - \mathbf{r})$, where $\mathbf{0} \leqslant \mathbf{E}(\mathbf{r}) \leqslant \mathbf{M}$, and $\mathbf{E}(\mathbf{r}) = O(\|\mathbf{1} - \mathbf{r}\|)$ as $\mathbf{r} \to \mathbf{1}$. Then $\rho^{-1}\mathbf{E}(\mathbf{f}_{k-1}(\mathbf{0})) = \rho^{-1}O(\|\mathbf{1} - \mathbf{f}_{k-1}(\mathbf{0})\|) = O(\rho^{k-2})$, which implies the existence of some constant $b > 0$ such that, for all $k \in \mathbb{N}$, $\mathbf{0} \leqslant \rho^{-1}\mathbf{E}(\mathbf{f}_{k-1}(\mathbf{0})) \leqslant \alpha_k \mathbf{R}$, with $\alpha_k := b\rho^{k-2}$. We thus have provided an explicit formula for the sequences $(\delta_k)_k$ and $(\alpha_k)_k$ introduced by Joffe and Spitzer in [8]. Finally let us apply (2.26) to $n = \lfloor k/2 \rfloor$ and replace in this inequality δ_n and α_n by their explicit expressions that we got. We obtain that, for all $k \in \mathbb{N}$,

$$\left\| \frac{\mathbf{1} - \mathbf{f}_k(\mathbf{0})}{\mathbf{v} \cdot (\mathbf{1} - \mathbf{f}_k(\mathbf{0}))} - \mathbf{u} \right\| \leqslant \frac{2^{2-s}ak^{s-1}(|\lambda|/\rho)^{k/2} + (b/\rho(1-\rho))\rho^{k/2}}{1 - 2^{1-s}ak^{s-1}(|\lambda|/\rho)^{k/2} - (b/\rho(1-\rho))\rho^{k/2}}. \tag{2.27}$$

Consequently, the right member of (2.24) will decrease with $\max\{k^{s-1}(|\lambda|\rho^{-1})^{k/2}, b(\rho(1 - \rho))^{-1}\rho^{k/2}\}$, as $k \to \infty$. $\square$

Hence the concept of the Q-process will have most practical relevance to approximate the very late extinction case if ρ is near to zero and if $|\lambda|$ is small compared with ρ. Note however that the very late extinction scenario is more likely to happen if ρ is near to unity because the time to extinction in most realizations will then be long (see Section 2.1).

2.5. Estimation of the Infection Parameter

We assume for this subsection that the parameters Ψ_k of the epidemic model (1.1)–(2.1) are not entirely known. More precisely, we assume that the Ψ_k are of the form, for all $k = 1 \cdots d$,

$$\Psi_k(\theta_0) = a_k\theta_0 + b_k, \tag{2.28}$$

where $a_k > 0$ and $b_k \geqslant 0$ are constants, and θ_0 is an unknown real parameter. We will write in what follows $\mathbf{\Psi}(\theta_0) = \mathbf{a}\theta_0 + \mathbf{b}$, where $\mathbf{\Psi}(\theta_0) := (\Psi_1(\theta_0), \ldots, \Psi_d(\theta_0))$ and so forth. This general assumption corresponds in particular to the case where the Ψ_k are of the form (1.2).

We estimate θ_0 by the following WCLSE in model (1.1)–(2.1). This estimator generalizes the well-known Harris estimator [12] in a BGW process. Let $\Theta :=]\theta_1, \theta_2[$, $\theta_2 > \theta_1 > 0$, such that $\theta_0 \in \Theta$. The WCLSE is based on the normalized process $Y_n := X_n/\sqrt{\mathbf{a} \cdot \mathbf{X}_{n-1}}$ and is defined by

$$\widehat{\theta}_{|\mathbf{X}_0|} := \arg\min_{\theta \in \Theta} \sum_{k=1}^{n} (Y_k - \mathbb{E}_\theta(Y_k \mid \mathbf{X}_{k-1}))^2 = \arg\min_{\theta \in \Theta} \sum_{k=1}^{n} \frac{(X_k - \mathbf{\Psi}(\theta) \cdot \mathbf{X}_{k-1})^2}{\mathbf{a} \cdot \mathbf{X}_{k-1}}. \tag{2.29}$$

We easily derive the following explicit form:

$$\widehat{\theta}_{|\mathbf{X}_0|} = \frac{\sum_{k=1}^{n}(X_k - \mathbf{b} \cdot \mathbf{X}_{k-1})}{\sum_{k=1}^{n} \mathbf{a} \cdot \mathbf{X}_{k-1}}. \tag{2.30}$$

The normalization of the process X_n by $\sqrt{\mathbf{a} \cdot \mathbf{X}_{n-1}}$ appears to be the most natural and suitable for the following reasons. First, this normalization generalizes the normalization $X_n/\sqrt{aX_{n-1}}$ in the monotype case, which is the one leading to the Harris estimator $\hat{m}$ of $m_0 = a\theta_0 + b$ since we have, for $d = 1$, $a\hat{\theta}_{X_0} + b = \hat{m}$. It also corresponds, in the linear case $\mathbf{b} = \mathbf{0}$, to the maximum likelihood estimator of θ_0. In addition, defining for any vector $\mathbf{x}$, $\underline{x} := \min_i x_i$ and $\overline{x} := \max_i x_i$, we have

$$\theta_0 + \frac{\underline{\mathbf{b}}}{\overline{\mathbf{a}}} \leqslant \mathbb{E}_{\theta_0}\left(\left(Y_k - \mathbb{E}_{\theta_0}(Y_k \mid \mathbf{X}_{k-1})\right)^2 \mid \mathbf{X}_{k-1}\right) = \theta_0 + \frac{\mathbf{b} \cdot \mathbf{X}_{k-1}}{\mathbf{a} \cdot \mathbf{X}_{k-1}} \leqslant \theta_0 + \frac{\overline{\mathbf{b}}}{\underline{\mathbf{a}}}, \tag{2.31}$$

hence the conditional variance of the error term $Y_k - \mathbb{E}_{\theta_0}(Y_k \mid \mathbf{X}_{k-1})$ in the stochastic regression equation $Y_k = \mathbb{E}_{\theta_0}(Y_k \mid \mathbf{X}_{k-1}) + Y_k - \mathbb{E}_{\theta_0}(Y_k \mid \mathbf{X}_{k-1})$ is invariant under multiplication of the whole process, and bounded respectively to $\{\mathbf{X}_n\}_n$, leading to the quasi-optimality of $\hat{\theta}_{|\mathbf{X}_0|}$ at finite $|\mathbf{X}_0|$ and n, in the sense of [13].

Let us provide asymptotic results for the estimator $\hat{\theta}_{|\mathbf{X}_0|}$ defined by (2.30), as the initial population size $|\mathbf{X}_0| = X_0 + X_{-1} + \cdots + X_{-d+1}$ tends to infinity. We denote by $m_{ij}^{(k)}(\theta)$ the (i,j)th entry in the kth power of the matrix $\mathbf{M}(\theta)$ given by (2.3).

Theorem 2.6. *Let us assume that, for each $i = 1 \cdots d$, there exists some $\alpha_i \in [0,1]$ such that $\lim_{|\mathbf{X}_0| \to \infty} X_{0,i}|\mathbf{X}_0|^{-1} = \alpha_i$. Then $\hat{\theta}_{|\mathbf{X}_0|}$ is strongly consistent, that is, $\lim_{|\mathbf{X}_0| \to \infty} \hat{\theta}_{|\mathbf{X}_0|} \overset{a.s.}{=} \theta_0$, and is asymptotically normally distributed:*

$$\lim_{|\mathbf{X}_0| \to \infty} \sqrt{\frac{\sum_{k=1}^n \mathbf{a} \cdot \mathbf{X}_{k-1}}{\sigma^2\left(\hat{\theta}_{|\mathbf{X}_0|}\right)}} \left(\hat{\theta}_{|\mathbf{X}_0|} - \theta_0\right) \overset{\mathscr{D}}{=} \mathcal{N}(0,1), \tag{2.32}$$

where

$$\sigma^2(\theta) := \theta + \frac{\sum_{k=1}^n \sum_{j=1}^d \sum_{i=1}^d \alpha_j b_i m_{ji}^{(k-1)}(\theta)}{\sum_{k=1}^n \sum_{j=1}^d \sum_{i=1}^d \alpha_j b_i m_{ji}^{(k-1)}(\theta)}. \tag{2.33}$$

Proof. Let us first prove that, for each $k = 1 \cdots n$ and each $i = 1 \cdots d$,

$$\lim_{|\mathbf{X}_0| \to \infty} \frac{X_{k,i}}{|\mathbf{X}_0|} \overset{a.s.}{=} \sum_{j=1}^d \alpha_j m_{ji}^{(k)}(\theta_0). \tag{2.34}$$

Using the branching property of the process $\{\mathbf{X}_n\}_n$ ([6]) we write

$$X_{k,i} = \sum_{j=1}^{X_{0,1}} X_{k,i,j}^{(1)} + \cdots + \sum_{j=1}^{X_{0,d}} X_{k,i,j}^{(d)}, \tag{2.35}$$

where, for all $l = 1 \cdots d$ and $j = 1 \cdots X_{0,l}$, $X_{k,i,j}^{(l)}$ is the ith coordinate of a d-type branching process at time k initialized by a single particle of type l. For k, i, and l fixed the random

variables $\{X_{k,i,j}^{(l)}\}_j$ are i.i.d. with mean value $m_{li}^{(k)}(\theta_0)$. According to the strong law of large numbers and under the theorem assumption, we have, for every $l = 1 \cdots d$ such that $X_{0,l} \neq 0$,

$$\lim_{|\mathbf{x}_0| \to \infty} \frac{\sum_{j=1}^{X_{0,l}} X_{k,i,j}^{(l)}}{X_{0,l}} \overset{\text{a.s.}}{=} m_{li}^{(k)}(\theta_0), \tag{2.36}$$

which together with the theorem assumption leads to (2.34).

To prove the consistency of $\widehat{\theta}_{|\mathbf{x}_0|}$ we apply (2.34) to (2.30), using the fact that $X_k = X_{k,1}$ and $X_{k-i} = X_{k-1,i}$, and obtain

$$\lim_{|\mathbf{x}_0| \to \infty} \widehat{\theta}_{|\mathbf{x}_0|} \overset{\text{a.s.}}{=} \frac{\sum_{k=1}^{n} \sum_{j=1}^{d} \alpha_j \left(m_{j1}^{(k)}(\theta_0) - \sum_{i=1}^{d} b_i m_{ji}^{(k-1)}(\theta_0) \right)}{\sum_{k=1}^{n} \sum_{i=1}^{d} \sum_{j=1}^{d} a_i \alpha_j m_{ji}^{(k-1)}(\theta_0)}. \tag{2.37}$$

By definition,

$$m_{j1}^{(k)}(\theta_0) = \sum_{i=1}^{d} m_{ji}^{(k-1)}(\theta_0) m_{i1}(\theta_0) = \sum_{i=1}^{d} m_{ji}^{(k-1)}(\theta_0)(a_i \theta_0 + b_i), \tag{2.38}$$

hence (2.37) immediately leads to the strong consistency.

We are now interested in the asymptotic distribution of $\widehat{\theta}_{|\mathbf{x}_0|} - \theta_0$. We derive from (2.30) that

$$\sqrt{\sum_{k=1}^{n} \mathbf{a} \cdot \mathbf{X}_{k-1}} \left(\widehat{\theta}_{|\mathbf{x}_0|} - \theta_0 \right) = \frac{\sum_{k=1}^{n} (X_k - \mathbf{\Psi}(\theta_0) \cdot \mathbf{X}_{k-1})}{\sqrt{\sum_{k=1}^{n} \mathbf{a} \cdot \mathbf{X}_{k-1}}}. \tag{2.39}$$

By (1.1),

$$X_k - \mathbf{\Psi}(\theta_0) \cdot \mathbf{X}_{k-1} = \sum_{i=1}^{d} \sum_{j=1}^{X_{k-i}} (Y_{k-i,k,j} - \Psi_i(\theta_0)) =: \sum_{i=1}^{d} \sum_{j=1}^{X_{k-i}} \overset{\circ}{Y}_{k-i,k,j}, \tag{2.40}$$

where the $\{Y_{k-i,k,j}\}_j$ are i.i.d. given $\mathcal{F}_{k-1}$, following a Poisson distribution with parameter $\Psi_i(\theta_0)$, and the $\{Y_{k-i,k,j}\}_{i,j}$ are independent given $\mathcal{F}_{k-1}$. Renumbering the $\overset{\circ}{Y}_{k-i,k,j}$ we then obtain

$$\sum_{k=1}^{n} (X_k - \mathbf{\Psi}(\theta_0) \cdot \mathbf{X}_{k-1}) = \sum_{i=1}^{d} \sum_{j=1}^{\sum_{k=1}^{n} X_{k-i}} \overset{\circ}{Y}_{k-i,k,j}. \tag{2.41}$$

Applying a central limit theorem for the sum of a random number of independent random variables (see, e.g., [14]), we obtain that, for all $i = 1 \cdots d$,

$$\lim_{|\mathbf{X}_0| \to \infty} \frac{\sum_{j=1}^{\sum_{k=1}^{n} X_{k-i}} \overset{\circ}{Y}_{k-i,k,j}}{\sqrt{\sum_{k=1}^{n} X_{k-i}}} \overset{\mathcal{D}}{=} \mathcal{N}(0, a_i \theta_0 + b_i). \tag{2.42}$$

We have used the fact that $|\mathbf{X}_0|$ is a real positive sequence growing to infinity, and $\sum_{k=1}^{n} X_{k-i}$ is a sequence of integer-valued random variables such that $\sum_{k=1}^{n} X_{k-i}/|\mathbf{X}_0|$ converges in probability to a finite random variable. In our case the limit is actually deterministic, since we have shown in (2.34) that

$$\lim_{|\mathbf{X}_0| \to \infty} \frac{\sum_{k=1}^{n} X_{k-i}}{|\mathbf{X}_0|} \overset{a.s.}{=} \sum_{k=1}^{n} \sum_{j=1}^{d} \alpha_j m_{ji}^{(k-1)}(\theta_0). \tag{2.43}$$

Using (2.41) in (2.39), we write

$$\sqrt{\sum_{k=1}^{n} \mathbf{a} \cdot \mathbf{X}_{k-1}} \left(\widehat{\theta}_{|\mathbf{X}_0|} - \theta_0 \right) = \sum_{i=1}^{d} \frac{\sum_{j=1}^{\sum_{k=1}^{n} X_{k-i}} \overset{\circ}{\zeta}_{k-i,k,j}}{\sqrt{\sum_{k=1}^{n} X_{k-i}}} \frac{\sqrt{\sum_{k=1}^{n} X_{k-i}}}{\sqrt{\sum_{k=1}^{n} \mathbf{a} \cdot \mathbf{X}_{k-1}}}. \tag{2.44}$$

Using again (2.34),

$$\lim_{|\mathbf{X}_0| \to \infty} \frac{\sqrt{\sum_{k=1}^{n} \mathbf{a} \cdot \mathbf{X}_{k-i}}}{\sqrt{\sum_{k=1}^{n} \mathbf{a} \cdot \mathbf{X}_{k-1}}} \overset{a.s.}{=} \sqrt{\frac{\sum_{k=1}^{n} \sum_{j=1}^{d} \alpha_j m_{ji}^{(k-1)}(\theta_0)}{\sum_{k=1}^{n} \sum_{j=1}^{d} \sum_{l=1}^{d} \alpha_j a_l m_{jl}^{(k-1)}(\theta_0)}}, \tag{2.45}$$

which, combined to (2.42) and (2.44), implies by Slutsky's theorem that

$$\lim_{|\mathbf{X}_0| \to \infty} \sqrt{\sum_{k=1}^{n} \mathbf{a} \cdot \mathbf{X}_{k-1}} \left(\widehat{\theta}_{|\mathbf{X}_0|} - \theta_0 \right) \overset{\mathcal{D}}{=} \mathcal{N}\left(0, \sigma^2(\theta_0)\right). \tag{2.46}$$

By (2.33) and the strong consistency, $\lim_{|\mathbf{X}_0|} \sqrt{\sigma^2(\theta_0)} / \sqrt{\sigma^2(\widehat{\theta}_{|\mathbf{X}_0|})} \overset{a.s.}{=} 1$, from which we finally deduce (2.32). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

2.6. Estimation of the Infection Parameter in the Worst-Case Scenario

In order to make predictions of the evolution of the epidemic in case of a very late extinction, that is, in order to make predictions of the behavior of the conditioned process $\{X_n^*\}_n$ introduced in Section 2.4, we need to estimate the parameter θ_0 in the setting of this conditioned process. We point out that θ_0 does not play the same role in the conditioned process $\{X_n^*\}_n$ and in the unconditioned process $\{X_n\}_n$, since, as shown in Proposition 2.2, this parameter interferes not only in the Poisson random variable but also in the Bernoulli

one. It would thus be irrelevant to estimate θ_0 with an estimator aimed for the unconditioned process, such as $\widehat{\theta}_{|\mathbf{X}_0|}$. Let us notice that, according to (2.16), the process $\{\mathbf{X}_n^*\}_n$ could be written as a multitype branching process with state dependent immigration. Because of this state-dependence, and since the parameter θ_0 acts in a nonlinear way in the immigration, the methods developed in estimation theory for branching processes with immigration (see, e.g., [15]) cannot be directly applied here. Similarly as in Section 2.5 we consider the WCLSE based on the process $Y_n^* := X_n^* / \sqrt{\mathbf{a} \cdot \mathbf{X}_{n-1}^*}$, namely,

$$\widehat{\theta}_n^* := \arg\min_{\theta \in \Theta} S_n(\theta), \qquad S_n(\theta) := \sum_{k=1}^{n} \left(Y_k^* - f(\theta, \mathbf{X}_{k-1}^*)\right)^2, \tag{2.47}$$

where Θ is defined in Section 2.5, and where

$$f(\theta_0, \mathbf{X}_{k-1}^*) := \mathbb{E}_{\theta_0}\left(Y_k^* \mid \mathbf{X}_{k-1}^*\right) = \frac{\mathbf{X}_{k-1}^* \cdot \mathbf{\Psi}(\theta_0) + p(\theta_0, \mathbf{X}_{k-1}^*)}{\sqrt{\mathbf{a} \cdot \mathbf{X}_{k-1}^*}}. \tag{2.48}$$

Let $\varepsilon_k^* := Y_k^* - f(\theta_0, \mathbf{X}_{k-1}^*)$ be the error term between the normalized process and its conditional expectation. We obtain that

$$g(\theta_0, \mathbf{X}_{k-1}^*) := \mathbb{E}_{\theta_0}\left((\varepsilon_k^*)^2 \mid \mathbf{X}_{k-1}^*\right) = \frac{\mathbf{X}_{k-1}^* \cdot \mathbf{\Psi}(\theta_0) + p(\theta_0, \mathbf{X}_{k-1}^*)\left(1 - p(\theta_0, \mathbf{X}_{k-1}^*)\right)}{\mathbf{a} \cdot \mathbf{X}_{k-1}^*}, \tag{2.49}$$

which implies

$$\theta_0 + \frac{\underline{\mathbf{b}}}{\overline{\mathbf{a}}} \leqslant \mathbb{E}_{\theta_0}\left((\varepsilon_k^*)^2 \mid \mathbf{X}_{k-1}^*\right) \leqslant \theta_0 + \frac{\overline{\mathbf{b}}+1}{\underline{\mathbf{a}}}. \tag{2.50}$$

In what follows, we denote by f' the derivative of f with respect to θ, and similarly for the other quantities depending on θ.

Theorem 2.7. *The estimator $\widehat{\theta}_n^*$ is strongly consistent, that is, $\lim_{n\to\infty} \widehat{\theta}_n^* \overset{a.s.}{=} \theta_0$, and has the following asymptotic distribution:*

$$\lim_{n \to \infty} \frac{\sum_{k=0}^{n} f'\left(\widehat{\theta}_n^*, \mathbf{X}_k^*\right)^2}{\sqrt{\sum_{k=0}^{n} f'\left(\widehat{\theta}_n^*, \mathbf{X}_k^*\right)^2 g\left(\widehat{\theta}_n^*, \mathbf{X}_k^*\right)}} \left(\widehat{\theta}_n^* - \theta_0\right) \overset{\mathcal{D}}{=} \mathcal{N}(0, 1). \tag{2.51}$$

Remark 2.8. Note that (2.51) involves the function f' and thus requires the knowledge of the derivative of the function u_j given in (2.5), which is not an explicit function of θ since ρ is not either. However, ρ' satisfies $\rho' = \sum_{k=1}^{d} a_k \rho^{-k} [\sum_{k=1}^{d} k(a_k\theta + b_k)\rho^{-k-1}]^{-1}$; hence u_j' is known as soon as ρ can be computed. Consequently, denoting ρ by $\rho(\theta)$ when θ is the parameter of the model, (2.51) can be used as soon as $\rho(\widehat{\theta}_n^{*\text{obs}})$ is known; for this purpose one can for instance numerically approximate the largest solution ρ of $\sum_{k=1}^{d} \Psi_k(\widehat{\theta}_n^{*\text{obs}})\rho^{-k} = 1$.

Proof. The proof heavily relies on a strong law of large numbers for homogeneous irreducible positive recurrent Markov chains applied to the conditioned process $\{\mathbf{X}_n^*\}_n$ and its stationary distribution π_{θ_0} (given by Proposition 2.4), which states [16]: for every π_{θ_0}-integrable function $h : \mathbb{N}^d \setminus \{\mathbf{0}\} \to \mathbb{R}$,

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} h(\mathbf{X}_k^*) \overset{\text{a.s.}}{=} \sum_{\mathbf{j} \in \mathbb{N}^d} h(\mathbf{j}) \pi_{\theta_0}(\mathbf{j}). \tag{2.52}$$

Note that the Perron's root $\rho(\theta)$ of $\mathbf{M}(\theta)$ and the associated right normalized eigenvector $\mathbf{u}(\theta)$ are $\mathcal{C}^\infty$-functions of θ.

In order to prove the strong consistency of $\widehat{\theta}_n^*$ as $n \to \infty$, since $f''(\cdot, \mathbf{X}_{k-1}^*)$ is not linear, we cannot use the general standard method based on the first-order expansion of $S_n'(\widehat{\theta}_n^*)$ at θ_0, and on the strong law of large numbers for martingales applied to $S_n'(\theta_0)$ correctly normalized. We consequently use the following conditions given in [17]:

(i) $f(., \mathbf{X}_{k-1}^*)$ is Lipschitz on Θ; that is, there exists a nonnegative $\mathcal{F}_{k-1}^*$-measurable function A_k (where $\mathcal{F}_{k-1}^* := \sigma(\mathbf{X}_0^*, \dots, \mathbf{X}_{k-1}^*)$), satisfying, for all $\delta_1, \delta_2 \in \Theta$, $|f(\delta_1, \mathbf{X}_{k-1}^*) - f(\delta_2, \mathbf{X}_{k-1}^*)| \leqslant A_k |\delta_1 - \delta_2|$ a.s.,

(ii) $\overline{\lim}_{k \to \infty} \mathbb{E}_{\theta_0}((\varepsilon_k^*)^2 \mid \mathbf{X}_{k-1}^*) \overset{\text{a.s.}}{<} \infty$,

(iii) $\lim_{n \to \infty} \inf_{\substack{\theta \in \Theta \\ |\theta - \theta_0| \geqslant \delta}} \sum_{k=1}^n (f(\theta_0, \mathbf{X}_{k-1}^*) - f(\theta, \mathbf{X}_{k-1}^*))^2 \overset{\text{a.s.}}{=} \infty$.

Let us note that, in the frame of a general model f, since $\overline{\Theta}$ is compact, (i) is satisfied as soon as $f(\theta, \mathbf{X}_{k-1}^*)$ has a first derivative in θ with $\sup_{\theta \in \Theta} f'(\theta, \mathbf{X}_{k-1}^*) < \infty$, (ii) is satisfied for any optimal estimator in the sense of [13] and moreover could be weakened (see [17]), and (iii) is a necessary condition. First, for all $\theta \in \Theta$ and $\mathbf{j} \in \mathbb{N}^d$, $\mathbf{j} \neq \mathbf{0}$, $f'(\theta, \mathbf{j}) = (\mathbf{a} \cdot \mathbf{j} + p'(\theta, \mathbf{j}))(\mathbf{a} \cdot \mathbf{j})^{-1/2}$, where p' denotes the derivative of p with respect to θ, which thanks to (2.17) is bounded on Θ. Condition (i) is thus satisfied. Condition (ii) follows from (2.50). Condition (iii) comes from the fact that, for every $\delta > 0$ and every $\theta \in \Theta$ such that $|\theta - \theta_0| \geqslant \delta$, applying the mean value theorem to the $\mathcal{C}^1$-function $p(\cdot, \mathbf{X}_{k-1}^*)$,

$$\sum_{k=1}^n (f(\theta_0, \mathbf{X}_{k-1}^*) - f(\theta, \mathbf{X}_{k-1}^*))^2 = (\theta_0 - \theta)^2 \sum_{k=1}^n \mathbf{a} \cdot \mathbf{X}_{k-1}^* \left(1 + \frac{p(\theta_0, \mathbf{X}_{k-1}^*) - p(\theta, \mathbf{X}_{k-1}^*)}{(\theta_0 - \theta) \mathbf{a} \cdot \mathbf{X}_{k-1}^*} \right)^2$$

$$\geqslant \delta^2 \sum_{k=1}^n \mathbf{a} \cdot \mathbf{X}_{k-1}^* \inf_{\theta \in \overline{\Theta}} \left(1 + \frac{p'(\theta, \mathbf{X}_{k-1}^*)}{\mathbf{a} \cdot \mathbf{X}_{k-1}^*} \right)^2. \tag{2.53}$$

Let us show that the function $\mathbf{j} \mapsto \mathbf{a} \cdot \mathbf{j} \inf_{\theta \in \overline{\Theta}}(1 + p'(\theta, \mathbf{j})/\mathbf{a} \cdot \mathbf{j})^2$ is π_{θ_0}-integrable. For every $\theta \in \overline{\Theta}$, $\mathbf{j} \in \mathbb{N}^d$ and $\mathbf{j} \neq \mathbf{0}$, denoting $\Xi := \sup_{i,\theta \in \overline{\Theta}}\{u_i(\theta), |\, u_i'(\theta)\,|\} < \infty$ (by continuity of $u_i'(\theta)$ on $\Theta_\gamma \supset \overline{\Theta}$), and $\underline{\mathbf{u}} := \min_{i,\theta \in \overline{\Theta}} u_i(\theta) > 0$,

$$
\begin{aligned}
\left| p'^{(\theta,\mathbf{j})} \right| &= \left| \frac{(u_i'(\theta)\mathbf{j} \cdot \mathbf{\Psi}(\theta) + u_1(\theta)\mathbf{j} \cdot \mathbf{a}) \sum_{i=2}^d j_i u_i(\theta) - u_1(\theta)\mathbf{j} \cdot \mathbf{\Psi}(\theta) \sum_{i=2}^d j_i u_i'(\theta)}{\left(u_1(\theta)\mathbf{j} \cdot \mathbf{\Psi}(\theta) + \sum_{i=2}^d j_i u_i(\theta)\right)^2} \right| \\
&\leqslant \frac{\Xi^2}{\underline{\mathbf{u}}^2} \frac{3\mathbf{j} \cdot \mathbf{\Psi}(\theta) \sum_{i=2}^d j_i}{\left(\mathbf{j} \cdot \mathbf{\Psi}(\theta) + \sum_{i=2}^d j_i\right)^2} \leqslant \frac{3\Xi^2}{4\underline{\mathbf{u}}^2} =: C_1.
\end{aligned}
\tag{2.54}
$$

Hence, for all $\mathbf{j} \neq \mathbf{0}$, $|\, \mathbf{a} \cdot \mathbf{j} \inf_{\theta \in \overline{\Theta}}(1 + p'(\theta, \mathbf{j})/\mathbf{a} \cdot \mathbf{j})^2 \,| \leqslant (1 + C_1/\underline{\mathbf{a}})^2 \mathbf{a} \cdot \mathbf{j}$, and applying (2.52) together with (2.22) we obtain that

$$
\lim_{n \to \infty} \frac{1}{n+1} \sum_{k=1}^n \mathbf{a} \cdot \mathbf{X}_{k-1}^* \inf_{\theta \in \overline{\Theta}}\left(1 + \frac{p'(\theta, \mathbf{X}_{k-1}^*)}{\mathbf{a} \cdot \mathbf{X}_{k-1}^*}\right)^2 \overset{\text{a.s.}}{=} \sum_{\mathbf{j} \in \mathbb{N}^d} \mathbf{a} \cdot \mathbf{j} \inf_{\theta \in \overline{\Theta}}\left(1 + \frac{p'(\theta, \mathbf{j})}{\mathbf{a} \cdot \mathbf{j}}\right)^2 \pi_{\theta_0}(\mathbf{j}).
\tag{2.55}
$$

Let $\mathbf{j} \neq \mathbf{0}$ fixed. Since, for all $\theta \in \overline{\Theta}$, $p'(\theta, \mathbf{j}) \neq -\mathbf{a} \cdot \mathbf{j}$, the extreme value theorem implies that $\inf_{\theta \in \overline{\Theta}}(1 + p'(\theta, \mathbf{j})/\mathbf{a} \cdot \mathbf{j})^2 > 0$. Hence the right term in (2.55) is strictly positive, which together with (2.55) leads to (iii).

Let us now consider the asymptotic distribution of $\widehat{\theta}_n^* - \theta_0$. For this purpose, we follow the steps of the proof of Proposition 6.1 in [17]. Writing the Taylor expansion of $S_n'(\widehat{\theta}_n^*)$ at θ_0 we obtain that $\widehat{\theta}_n^* - \theta_0 = -S_n'(\theta_0)/S_n''(\widetilde{\theta}_n)$, for some $\widetilde{\theta}_n = \theta_0 + t_n(\widehat{\theta}_n^* - \theta_0)$, with $t_n \in \,]0,1[$. Since $S_n'(\theta_0) = -2\sum_{k=1}^n \varepsilon_k^* f'(\theta_0, \mathbf{X}_{k-1}^*)$, we can write

$$
\sqrt{n}\left(\widehat{\theta}_n^* - \theta_0\right) = \frac{\sum_{k=1}^n \varepsilon_k^* f'(\theta_0, \mathbf{X}_{k-1}^*)}{\sqrt{n}} \left(\frac{F_n}{n}\right)^{-1} \left(\frac{1}{2} \frac{S_n''(\widetilde{\theta}_n)}{F_n}\right)^{-1},
\tag{2.56}
$$

where $F_n := \sum_{k=1}^n (f'(\theta_0, \mathbf{X}_{k-1}^*))^2$. Let us first show that

$$
\lim_{n \to \infty} \frac{F_n}{n} \overset{\text{a.s.}}{=} \sum_{\mathbf{j} \in \mathbb{N}^d} (f'(\theta_0, \mathbf{j}))^2 \pi_{\theta_0}(\mathbf{j}).
\tag{2.57}
$$

This is an application of (2.52) and (2.22), since, for all $\mathbf{j} \in \mathbb{N}^d$, $\mathbf{j} \neq \mathbf{0}$,

$$
(f'(\theta_0, \mathbf{j}))^2 \leqslant \mathbf{a} \cdot \mathbf{j} + 2C_1 + \frac{C_1^2}{\underline{\mathbf{a}}}.
\tag{2.58}
$$

In view of (2.56), we now prove that

$$\lim_{n \to \infty} \frac{S_n''\left(\tilde{\theta}_n\right)}{F_n} \overset{\text{a.s.}}{=} 2. \tag{2.59}$$

Computing S_n'' thanks to the formula $S_n(\theta) = \sum_{k=1}^n \left(\varepsilon_k^* + f(\theta_0, \mathbf{X}_{k-1}^*) - f(\theta, \mathbf{X}_{k-1}^*)\right)^2$, it appears that (2.59) is true, as soon as the following holds:

$$\limsup_{\substack{n \to \infty \\ \theta \in \Theta}} \frac{\left|\sum_{k=1}^n \varepsilon_k^* f''(\theta, \mathbf{X}_{k-1}^*)\right|}{F_n} \overset{\text{a.s.}}{=} 0, \tag{2.60}$$

$$\lim_{n \to \infty} \frac{\sum_{k=1}^n \left(f'\left(\tilde{\theta}_n, \mathbf{X}_{k-1}^*\right)\right)^2}{F_n} \overset{\text{a.s.}}{=} 1, \tag{2.61}$$

$$\lim_{n \to \infty} \frac{\sum_{k=1}^n \left(f(\theta_0, \mathbf{X}_{k-1}^*) - f\left(\tilde{\theta}_n, \mathbf{X}_{k-1}^*\right)\right) f''\left(\tilde{\theta}_n, \mathbf{X}_{k-1}^*\right)}{F_n} \overset{\text{a.s.}}{=} 0. \tag{2.62}$$

Let us prove (2.60)–(2.62). First, (2.60) is given by a strong law of large numbers proved in [17, Proposition 5.1]. The latter can be indeed applied since $\lim_n F_n \overset{\text{a.s.}}{=} \infty$ (as an immediate consequence of the stronger result (2.57)), and since $f''(\cdot, \mathbf{X}_{k-1}^*)$ fulfills the required Lipschitz condition. Indeed we proved earlier that $u_i'''(\theta)$ is continuous on the compact set $\overline{\Theta}$ and is thus bounded on $\overline{\Theta}$, which implies that $f'''(\cdot, \mathbf{X}_{k-1}^*) = p'''(\cdot, \mathbf{X}_{k-1}^*)(\mathbf{a} \cdot \mathbf{X}_{k-1}^*)^{-1/2}$ is bounded by a $\mathcal{F}_{k-1}^*$-measurable function. In view of (2.61), we consider the function $f(\theta, \mathbf{j})^2$ and its derivative $2f'(\theta, \mathbf{j}) f''(\theta, \mathbf{j})$. Similarly as for (2.54), one can show that there exists a constant $C_2 > 0$ such that for all $\theta \in \overline{\Theta}$, and all $\mathbf{j} \neq \mathbf{0}$, $|p''(\theta, \mathbf{j})| \leqslant C_2$. This implies

$$\left|2f'(\theta, \mathbf{j}) f''(\theta, \mathbf{j})\right| \leqslant 2\left|\frac{(\mathbf{a} \cdot \mathbf{j} + p'(\theta, \mathbf{j})) p''(\theta, \mathbf{j})}{\mathbf{a} \cdot \mathbf{j}}\right| \leqslant 2C_2\left(1 + \frac{C_1}{\underline{\mathbf{a}}}\right). \tag{2.63}$$

Consequently,

$$\frac{\left|\sum_{k=1}^n \left(f'\left(\tilde{\theta}_n, \mathbf{X}_{k-1}^*\right)\right)^2 - (f'(\theta_0, \mathbf{X}_{k-1}^*))^2\right|}{F_n} \leqslant 2C_2\left(1 + \frac{C_1}{\underline{\mathbf{a}}}\right)\left|\hat{\theta}_n^* - \theta_0\right|\left(\frac{F_n}{n}\right)^{-1}, \tag{2.64}$$

which by (2.57) and the strong consistency of $\hat{\theta}_n^*$ almost surely tends to 0. Writing

$$\frac{\sum_{k=1}^n f\left(\tilde{\theta}_n, \mathbf{X}_{k-1}^*\right)^2}{F_n} = 1 + \frac{\sum_{k=1}^n \left(\left(f'\left(\tilde{\theta}_n, \mathbf{X}_{k-1}^*\right)\right)^2 - (f'(\theta_0, \mathbf{X}_{k-1}^*))^2\right)}{F_n}, \tag{2.65}$$

this implies (2.61). It now remains to prove (2.62). We write

$$
\frac{\left|\sum_{k=1}^{n}\left(f(\theta_0, \mathbf{X}_{k-1}^*) - f(\tilde{\theta}_n, \mathbf{X}_{k-1}^*)\right) f''(\tilde{\theta}_n, \mathbf{X}_{k-1}^*)\right|}{F_n}
$$

$$
\leqslant \frac{1}{F_n} \sum_{k=1}^{n} \frac{\mathbf{a} \cdot \mathbf{X}_{k-1}^* \left|\theta_0 - \tilde{\theta}_n\right| + \left|p(\theta_0, \mathbf{X}_{k-1}^*) - p(\tilde{\theta}_n, \mathbf{X}_{k-1}^*)\right|}{\mathbf{a} \cdot \mathbf{X}_{k-1}^*} \left|p''(\tilde{\theta}_n, \mathbf{X}_{k-1}^*)\right| \qquad (2.66)
$$

$$
\leqslant \frac{1}{F_n} \sum_{k=1}^{n} \left(\left|\theta_0 - \tilde{\theta}_n\right| + \frac{C_1 \left|\theta_0 - \tilde{\theta}_n\right|}{\underline{a}} \right) C_2 \leqslant \left|\theta_0 - \hat{\theta}_n^*\right| C_2 \left(1 + \frac{C_1}{\underline{a}}\right) \left(\frac{F_n}{n}\right)^{-1},
$$

which thanks to (2.57) and the strong consistency of $\hat{\theta}_n^*$ implies (2.62). In view of (2.56), we finally want to prove that $\sum_{k=1}^{n} \varepsilon_k^* f'(\theta_0, \mathbf{X}_{k-1}^*)/\sqrt{n}$ converges in distribution. For this purpose we make use of a central limit theorem for martingale difference arrays [18, 19]; if $\{M_k^{(n)}, \mathcal{F}_k^{(n)}, 1 \leqslant k \leqslant n\}, n \geqslant 1$ is a sequence of square integrable martingales with associated Meyer process $\langle M \rangle^{(n)} = (\langle M \rangle_k^{(n)})_{1 \leqslant k \leqslant n}$ satisfying $\lim_{n\to\infty} \langle M_n \rangle^{(n)} \overset{P}{=} c^2$ for some constant c, and such that, for all $\varepsilon > 0$,

$$
\lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{E}\left[\left|M_k^{(n)} - M_{k-1}^{(n)}\right|^2 \mathbf{1}_{\{|M_k^{(n)} - M_{k-1}^{(n)}| \geqslant \varepsilon\}} \mid \mathcal{F}_{k-1}^{(n)} \right] \overset{P}{=} 0, \qquad (2.67)
$$

then $\lim_{n\to\infty} M_n^{(n)} \overset{\mathfrak{D}}{=} \mathcal{N}(0, c^2)$. Let us define $M_k^{(n)} := \sum_{l=1}^{k} \varepsilon_l^* f'(\theta_0, \mathbf{X}_{l-1}^*)/\sqrt{n}$, for every $k \leqslant n$. First, for any $k \leqslant n$, $\mathbb{E}_{\theta_0}(\varepsilon_k^* f'(\theta_0, \mathbf{X}_{k-1}^*)/\sqrt{n} \mid \mathbf{X}_{k-1}^*) = 0$. Second,

$$
\mathbb{E}_{\theta_0}\left(\left(\frac{\varepsilon_k^* f'(\theta_0, \mathbf{X}_{k-1}^*)}{\sqrt{n}} \right)^2 \mid \mathbf{X}_{k-1}^* \right) = \frac{(f'(\theta_0, \mathbf{X}_{k-1}^*))^2 g(\theta_0, \mathbf{X}_{k-1}^*)}{n}, \qquad (2.68)
$$

hence $\{M_k^{(n)}\}_{k \leqslant n}$ is a sequence of square integrable martingales. Moreover, using inequalities (2.50) and (2.58), we obtain, by (2.22),

$$
\sum_{\mathbf{j} \in \mathbb{N}^d} (f'(\theta_0, \mathbf{j}))^2 g(\theta_0, \mathbf{j}) \pi_{\theta_0}(\mathbf{j}) \leqslant \left(\theta_0 + \frac{\overline{\mathbf{b}} + 1}{\underline{a}} \right) \left(\sum_{\mathbf{j} \in \mathbb{N}^d} \mathbf{a} \cdot \mathbf{j} \pi_{\theta_0}(\mathbf{j}) + 2C_1 + \frac{C_1^2}{\underline{a}} \right) < \infty. \qquad (2.69)
$$

So, by means of (2.52),

$$
\lim_{n\to\infty} \langle M_n \rangle^{(n)} = \lim_{n\to\infty} \sum_{k=1}^{n} \mathbb{E}_{\theta_0}\left(\left(\frac{\varepsilon_k f'(\theta_0, \mathbf{X}_{k-1}^*)}{\sqrt{n}} \right)^2 \mid \mathbf{X}_{k-1}^* \right)
$$

$$
\overset{a.s.}{=} \sum_{\mathbf{j} \in \mathbb{N}^d} (f'(\theta_0, \mathbf{j}))^2 g(\theta_0, \mathbf{j}) \pi_{\theta_0}(\mathbf{j}). \qquad (2.70)
$$

Third, using Cauchy-Schwarz and Bienaymé-Chebyshev inequalities,

$$
\sum_{k=1}^{n} \mathbb{E}_{\theta_0} \left[\left| \frac{\varepsilon_k^* f'(\theta_0, \mathbf{X}_{k-1}^*)}{\sqrt{n}} \right|^2 \mathbf{1}_{\{|\varepsilon_k^* f'(\theta_0, \mathbf{X}_{k-1}^*)/\sqrt{n}| \geqslant \varepsilon\}} \mid \mathbf{X}_{k-1}^* \right]
$$
$$
\leqslant \sum_{k=1}^{n} \left(\mathbb{E}_{\theta_0} \left[\left| \frac{\varepsilon_k^* f'(\theta_0, \mathbf{X}_{k-1}^*)}{\sqrt{n}} \right|^4 \mid \mathbf{X}_{k-1}^* \right] \right)^{1/2} \left(\mathbb{P}_{\theta_0} \left[\left| \frac{\varepsilon_k^* f'(\theta_0, \mathbf{X}_{k-1}^*)}{\sqrt{n}} \right| \geqslant \varepsilon \mid \mathbf{X}_{k-1}^* \right] \right)^{1/2}
$$
$$
\leqslant \frac{1}{n^{3/2}\varepsilon} \sum_{k=1}^{n} |f'(\theta_0, \mathbf{X}_{k-1}^*)|^3 \left(\mathbb{E}_{\theta_0} \left[(\varepsilon_k^*)^4 \mid \mathbf{X}_{k-1}^* \right] \right)^{1/2} \left(\mathbb{E}_{\theta_0} \left[(\varepsilon_k^*)^2 \mid \mathbf{X}_{k-1}^* \right] \right)^{1/2}.
$$

$$(2.71)$$

Let us compute $\mathbb{E}_{\theta_0}((\varepsilon_k^*)^4 \mid \mathbf{X}_{k-1}^*)$. We can show that the 4th central moment of the independent sum of a Poisson and a Bernoulli random variables equals $\mu_4 + 6\mu_2\gamma_2 + \gamma_4$, where μ_i and γ_i denote the ith central moment of the Poisson and of the Bernoulli variable. If these variables have parameter λ and p, respectively, then $\mu_4 = \lambda(1 + 3\lambda)$, $\mu_2 = \lambda$, $\gamma_4 = p(1-p)(3p^2 - 3p + 1) \in [0,1]$, and $\gamma_2 = p(1-p) \in [0,1]$. We thus obtain

$$
\left| \mathbb{E}_{\theta_0} \left((\varepsilon_k^*)^4 \mid \mathbf{X}_{k-1}^* \right) \right| \leqslant \frac{\mathbf{\Psi}(\theta_0) \cdot \mathbf{X}_{k-1}^* (7 + 3\mathbf{\Psi}(\theta_0) \cdot \mathbf{X}_{k-1}^*) + 1}{(\mathbf{a} \cdot \mathbf{X}_{k-1}^*)^2}.
$$

$$(2.72)$$

Hence

$$
\left| f^{*'}(\theta_0, \mathbf{X}_{k-1}^*) \right|^3 \left(\mathbb{E}_{\theta_0} \left[\varepsilon_k^4 \mid \mathbf{X}_{k-1}^* \right] \right)^{1/2} \left(\mathbb{E}_{\theta_0} \left[\varepsilon_k^2 \mid \mathbf{X}_{k-1}^* \right] \right)^{1/2}
$$
$$
\leqslant \left(\sqrt{\mathbf{a} \cdot \mathbf{X}_{k-1}^*} + \frac{C_1}{\sqrt{\underline{\mathbf{a}}}} \right)^3 \sqrt{\frac{\mathbf{\Psi}(\theta_0) \cdot \mathbf{X}_{k-1}^* (7 + 3\mathbf{\Psi}(\theta_0) \cdot \mathbf{X}_{k-1}^*) + 1}{\mathbf{a} \cdot \mathbf{X}_{k-1}^*}} \left(\theta_0 + \frac{\overline{\mathbf{b}} + 1}{\underline{\mathbf{a}}} \right)^{1/2}.
$$

$$(2.73)$$

Since the highest power of X_n^* involved in (2.73) is $3/2$, and since by Proposition 2.4 the stationary distribution π_{θ_0} has finite second moments, we can apply (2.52) to (2.71) and obtain that

$$
\lim_{n \to \infty} \sum_{k=1}^{n} \mathbb{E}_{\theta_0} \left[\left| \frac{\varepsilon_k^* f'(\theta_0, \mathbf{X}_{k-1}^*)}{\sqrt{n}} \right|^2 \mathbf{1}_{\{|\varepsilon_k^* f'(\theta_0, \mathbf{X}_{k-1}^*)/\sqrt{n}| \geqslant \varepsilon\}} \mid \mathbf{X}_{k-1}^* \right] \overset{\text{a.s.}}{=} 0.
$$

$$(2.74)$$

It then ensues from the central limit theorem mentioned above that

$$
\lim_{n \to \infty} \frac{\sum_{k=1}^{n} \varepsilon_k^* f'(\theta_0, \mathbf{X}_{k-1}^*)}{\sqrt{n}} \overset{\mathfrak{D}}{=} \mathcal{N} \left(0, \sum_{\mathbf{j} \in \mathbb{N}^d} (f'(\theta_0, \mathbf{j}))^2 g(\theta_0, \mathbf{j}) \pi_{\theta_0}(\mathbf{j}) \right).
$$

$$(2.75)$$

Finally, (2.56), together with (2.57), (2.59), (2.75), and Slutsky's theorem, implies that

$$\lim_{n \to \infty} \sqrt{n}\left(\widehat{\theta}_n^* - \theta_0\right) \overset{\mathcal{D}}{=} \mathcal{N}\left(0, \frac{\sum_{\mathbf{j} \in \mathbb{N}^d} \left(f'(\theta_0, \mathbf{j})\right)^2 g(\theta_0, \mathbf{j}) \pi_{\theta_0}(\mathbf{j})}{\left(\sum_{\mathbf{j} \in \mathbb{N}^d} \left(f'(\theta_0, \mathbf{j})\right)^2 \pi_{\theta_0}(\mathbf{j})\right)^2}\right), \tag{2.76}$$

which leads to

$$\lim_{n \to \infty} \sqrt{n \frac{\left(\sum_{\mathbf{j} \in \mathbb{N}^d} f'(\theta_0, \mathbf{j})^2 \pi_{\theta_0}(\mathbf{j})\right)^2}{\sum_{\mathbf{j} \in \mathbb{N}^d} f'(\theta_0, \mathbf{j})^2 g(\theta_0, \mathbf{j}) \pi_{\theta_0}(\mathbf{j})}} \left(\widehat{\theta}_n^* - \theta_0\right) \overset{\mathcal{D}}{=} \mathcal{N}(0, 1). \tag{2.77}$$

It now remains to prove the asymptotic distribution (2.51). Thanks to what precedes, this result is immediate as soon as we prove that

$$\lim_{n \to \infty} \frac{1}{n+1} \sum_{k=0}^{n} \left(f'\left(\widehat{\theta}_n^*, \mathbf{X}_k^*\right)\right)^2 g\left(\widehat{\theta}_n^*, \mathbf{X}_k^*\right) \overset{\text{a.s.}}{=} \sum_{\mathbf{j} \in \mathbb{N}^d} \left(f'(\theta_0, \mathbf{j})\right)^2 g(\theta_0, \mathbf{j}) \pi_{\theta_0}(\mathbf{j}), \tag{2.78}$$

as well as the equivalent result for the numerator. For this purpose, we write

$$\sum_{k=0}^{n} \left(f'\left(\widehat{\theta}_n^*, \mathbf{X}_k^*\right)\right)^2 g\left(\widehat{\theta}_n^*, \mathbf{X}_k^*\right) = \sum_{k=0}^{n} (f'(\theta_0, \mathbf{X}_k^*))^2 g(\theta_0, \mathbf{X}_k^*)$$

$$+ \sum_{k=0}^{n} \left[\left(f'\left(\widehat{\theta}_n^*, \mathbf{X}_k^*\right)\right)^2 g\left(\widehat{\theta}_n^*, \mathbf{X}_k^*\right) - (f'(\theta_0, \mathbf{X}_k^*))^2 g(\theta_0, \mathbf{X}_k^*)\right], \tag{2.79}$$

and show that $(f'(\cdot, \mathbf{j}))^2 g(\cdot, \mathbf{j})$ has a bounded derivative and is thus Lipschitz. We have indeed $| g'(\theta, \mathbf{j}) |=| \mathbf{a} \cdot \mathbf{j} + p'(\theta, \mathbf{j}) | (\mathbf{a} \cdot \mathbf{j})^{-1} \leqslant 1 + C_1 \underline{\mathbf{a}}^{-1}$, hence

$$\left|2 f''(\theta, \mathbf{j}) f'(\theta, \mathbf{j}) g(\theta, \mathbf{j}) + (f'(\theta, \mathbf{j}))^2 g'(\theta, \mathbf{j})\right|$$

$$\leqslant 2 C_2 \left(1 + C_1 \underline{\mathbf{a}}^{-1}\right) \underline{\mathbf{a}}^{-1} (\mathbf{\Psi}(\theta) \cdot \mathbf{j} + 1) + \left(1 + C_1 \underline{\mathbf{a}}^{-1}\right)^2 (\mathbf{a} \cdot \mathbf{j} + C_1) \leqslant C_3 \mathbf{\Psi}(\theta_2) \cdot \mathbf{j}, \tag{2.80}$$

for some constant $C_3 > 0$. This enables us to write

$$\frac{1}{n+1} \sum_{k=0}^{n} \left|\left(f'\left(\widehat{\theta}_n^*, \mathbf{X}_k^*\right)\right)^2 g\left(\widehat{\theta}_n^*, \mathbf{X}_k^*\right) - (f'(\theta_0, \mathbf{X}_k^*))^2 g(\theta_0, \mathbf{X}_k^*)\right|$$

$$\leqslant \left|\widehat{\theta}_n^* - \theta_0\right| C_3 \frac{1}{n+1} \sum_{k=0}^{n} \mathbf{\Psi}(\theta_2) \cdot \mathbf{X}_k^*. \tag{2.81}$$

By the strong consistency of $\widehat{\theta}_n^*$ together with (2.52) and (2.22), (2.81) almost surely tends to zero. Combined with (2.52) and (2.69) in (2.79), this implies (2.78). □

3. Application to Real Epidemics

3.1. Explicit Epidemic Model

Although the epidemic model (1.1) has a clear interpretation in terms of the disease propagation, it can be also proved [9] that it is the limit process, as the initial population size tends to infinity, of the incidence of infectives described by a thorough process in a large branching population, taking into account all the health states of the disease and describing in detail the infection and latent processes for each individual in the population. This detailed process is a multitype branching process with age and population size dependence, taking into account the variability of many individual factors such as the reproduction, the survival, the transmission of the disease, and the latent time. We assume a disease of the general SEIR type. We also assume a random latent period and, for the sake of simplicity, a duration of the state I of one time unit only. Hence the incidence of infectives exactly corresponds to the incidence of cases.

The main assumptions for the construction of the limit model $\{X_n\}_n$ are, concerning the disease as follows: (i) at the initial time the disease is rare and the total population size is large; (ii) the infection *via* horizontal route is of Reed-Frost type, with the probability for a susceptible individual to become infected by a given dose of pathogens being inversely proportional to the total population size; (iii) the individual survival law is the same for E and S individuals and is independent of the population and of the time; (iv) the latent time law given the individual survival is independent of the time, the individual age, and the infectives population during the latent period. This last property is possible if we assume that overinfection during the incubation has a negligible effect on the latent time. We moreover assume that the whole healthy population size is relatively stable over time.

The limit process (1.1) is then obtained in an inductive way as the limit in distribution of $\{I_n\}_n$ as the initial total population size $|\mathbf{N}_0| \to \infty$, where I_n denotes the number of new infectives at time n. More precisely, denoting by E_n the incidence of exposed individuals at time n and by a_m the largest individual survival age, we obtain that $\{X_n, Z_n\}_n \overset{\mathcal{D}}{=} \lim_{N_0 \to \infty} \{I_n, E_n\}_n$, with

$$X_n \mid (X_{n-1}, \ldots, X_{n-a_m+1}) \overset{\mathcal{D}}{=} \mathrm{Poiss}\left(\sum_{k=1}^{a_m-1} \Psi_k X_{n-k} \right), \tag{3.1}$$

$$Z_n \mid X_n \overset{\mathcal{D}}{=} \mathrm{Poiss}(\Psi_0 X_n), \tag{3.2}$$

and where for all $k = 1 \cdots a_m - 1$, Ψ_k satisfies (1.2) in Section 1, with the latent period distribution given survival, $P_{\mathrm{inc},a}$, assumed independent of a, being denoted by P_{inc}. Thus

$$\Psi_k = \theta_0 P_{\mathrm{inc}}(k) \sum_{a=k+1}^{a_m} P_{\mathrm{age}}(a) + p_{\mathrm{mat}} P_{\mathrm{inc}}(k) P_{\mathrm{age}}(k+1), \tag{3.3}$$

and similarly,

$$\Psi_0 = \theta_0 + p_{\mathrm{mat}} P_{\mathrm{age}}(1). \tag{3.4}$$

Moreover, when assuming a relatively stable population size and denoting by $S(a)$ the individual survival probability until age a at least, then $P_{\text{age}}(a)$ is equal to $S(a)(\sum_{a'=1}^{a_m} S(a'))^{-1}$.

3.2. Illustration: The BSE Epidemic in Great Britain

In this section we provide an illustration of the methodology developed in Section 2 by studying the decay phase of the BSE epidemic in Great Britain, based on the observations of the yearly diagnosed cases from 1989 until 2011, with 1988 being the date of the main control measure. The disease that was first officially identified in 1986 [20] reached its peak in 1992 (36682 cases) and is obviously now in its decay phase. Since only a very few cases were recently reported [5], namely, 11 in 2010 and 5 in 2011, the spread of the disease should *a priori* "soon" come to an end. It has been accepted that the epidemic is fading out, with a very low level of risk for cattle and humans. However it might be interesting to have a more precise idea of the extinction phase of the disease, for instance, of its length and of its intensity.

3.2.1. Choice of the Epidemic Model

BSE is a transmissible disease through ingestion of prions (horizontal route) and through maternal route. It is commonly accepted (see, e.g., [21]) that the infectious phase (including the clinical state) lasts at most one year, while the latent state can be of several years. If one compares the epidemic data with the total population of around 9 million cattle in Great Britain, it is reasonable to assume that, even at the peak time, BSE may be considered as a rare disease in a large population. We moreover assume that the probability for an animal to become infected by a horizontal route follows a Reed-Frost type model as in (ii) of Section 3.1, and that the probability for a calf to become infected by its dam is nonnull and constant over time. Moreover, for the sake of simplicity and parsimony, we do not take into account potential heterogeneity factors such as the different regulations from 1989 (the main regulation was the feed ban of July 1988 and was shown to be quite efficient [22]), the different types of breeding and of races, the age for animals older than one year, and the evolution of the surveillance system and of diagnosis tests. So the process $\{X_n\}_n$ defined by (3.1)–(3.4) appears relevant to model the spread of this fatal disease after the 1988 ban, where the infectious state I corresponds to the end of the incubation period and the clinical state, and R corresponds to the death (assumed to be either by routine slaughtering for E and S cattle, and control slaughtering for I cattle). Choosing a time step of one year, X_n then represents the yearly incidence of notified cases, which corresponds to the available data provided by the World Organisation for Animal Health [5].

3.2.2. Choice of the Model Parameters

In order to predict the future epidemic evolution, we need to evaluate all the parameters involved in (3.3). These parameters were already studied in previous works [22–25] using different models, amounts of observations, and kinds of estimators. Since it might have a crucial role in the decay phase of the disease, we focus here on θ_0 the infection parameter from 1989 *via* a horizontal route of transmission (i.e., the mean number per infective and per year of newly infected). Until the 1988 feed ban regulation, the main routes of transmission of BSE were horizontal *via* protein supplements (Meat and Bone Meal, milk replacers), and maternal from a dam to its calf. Since a previous statistical study [22] concluded to the full efficiency of the 1988 ban and since most of cattle are slaughtered before the age of 10 years,

the fact that cases of BSE are still observed more than 20 years later could suggest the existence of a remaining source of infection, either *via* a maternal transmission route, or *via* a horizontal one, for example, *via* the ingestion of excreted prions from alive infected animals, or from prions left in the environment. The prediction of the future disease spread thus strongly relies on the intensity of this infection, quantified by the parameter θ_0. We provide an estimation of θ_0 together with a confidence interval, which will be used for prediction estimations. We provide in addition a sensitivity analysis to the other parameters of the model, which are chosen as follows. First, the maximal age of the individuals in the population a_m is set to $a_m = 10$, which corresponds to the largest reported survival age with a nonnegligible probability. Next, we set $p_{\mathrm{mat}} = 0.1$, which is the largest order of magnitude commonly accepted for the maternal infection probability [23, 25]. Following a previous work, [22], we assume here a discretized Weibull distribution with parameters α, β for the distribution of the latent period; that is to say, for each $k \geqslant 1$, $P_{\mathrm{inc}}(k) := e^{-((\alpha-1)/\alpha\beta^\alpha)(k-1)^\alpha} - e^{-((\alpha-1)/\alpha\beta^\alpha)k^\alpha}$, where α is a shape parameter, and β is the mode of the probability density of the corresponding continuous Weibull distribution. The parameters α and β have a very bad identifiability with the infection parameters on a sole monotonous phase and are consequently estimated in [22] on the whole epidemic series (growth and decay), by Bayesian maximum *a posteriori* (MAP) estimations. We set $\alpha = \widehat{\alpha}^{\mathrm{MAP}} = 3.84$ and $\beta = \widehat{\beta}^{\mathrm{MAP}} = 7.46$. Since the death of susceptible and exposed animals is mostly due to routine slaughtering, the survival probability $S(k)$ (probability for an apparently healthy animal to survive at least until k years) corresponds to the probability of being slaughtered after the age of k years. It is derived from existing literature [24].

3.2.3. Estimation of the Infection Parameter

Our study is based on the yearly number of cases of BSE reported in Great Britain from 1989 until 2011 (Figure 1) provided by the World Organisation for Animal Health [5], denoted by X_n^{obs} for each year n. We set $d = a_m - 1 = 9$ and denote by $\mathbf{X}_n^{\mathrm{obs}}$ the d-dimensional vector $(X_n^{\mathrm{obs}}, \ldots, X_{n-d+1}^{\mathrm{obs}})$. We choose the first time $n = 1$ of the epidemic model such that the model with its initial values covers the period starting from 1989, which we consider here as a time-homogeneous period. Hence $X_{-d+1} = X_{1989}^{\mathrm{obs}}$, and $n = 1$ corresponds to the year $1989 + d = 1998$. We estimate θ_0 by the WCLSE (2.30) in model (1.1)–(3.3), where $\{p_{\mathrm{mat}}, \alpha, \beta, \{S(k)\}_k\}$ are given by the previous values, and where $\mathbf{X}_0 = \mathbf{X}_{1997}^{\mathrm{obs}}$. According to [5], $|\mathbf{X}_0| = |\mathbf{X}_{1997}^{\mathrm{obs}}| = 167977$. We are thus close to the asymptotic $|\mathbf{X}_0| \to \infty$. The number of observations is $n = 14$. The estimator (2.30) provides the estimation $\widehat{\theta}_{|\mathbf{X}_0|}^{\mathrm{obs}} = 2.4324$. We point out that this estimation is of the same order of magnitude as the maximum *a posteriori* Bayesian estimation $\widehat{\theta}^{\mathrm{MAP}} = 2.43$ based on the whole epidemic until 2007, assuming a uniform prior probability [22]. Using (2.32) we obtain the following confidence interval $[\widehat{\theta}_{\mathrm{min}}, \widehat{\theta}_{\mathrm{max}}]$ with asymptotic probability 95%, where $\widehat{\theta}_{\mathrm{min}} := \widehat{\theta}_{|\mathbf{X}_0|} - 1.96\widehat{c}_1^{-1}$, $\widehat{\theta}_{\mathrm{max}} := \widehat{\theta}_{|\mathbf{X}_0|} + 1.96\widehat{c}_1^{-1}$, and $\widehat{c}_1 := (\sum_{k=1}^n \mathbf{a} \cdot \mathbf{X}_{k-1}/\sigma^2(\widehat{\theta}_{|\mathbf{X}_0|}))^{1/2}$. Assuming $\alpha_i = X_{1997-i+1}^{\mathrm{obs}} |\mathbf{X}_{1997}^{\mathrm{obs}}|^{-1}$, we get $\widehat{c}_1^{\mathrm{obs}} = 40.7343$ (observed value of $\widehat{c}_1$), and

$$\mathbb{P}\left(\theta_0 \in \left[\widehat{\theta}_{\mathrm{min}}, \widehat{\theta}_{\mathrm{max}}\right]\right) \simeq 95\%, \qquad \widehat{\theta}_{\mathrm{min}}^{\mathrm{obs}} = 2.3842, \qquad \widehat{\theta}_{\mathrm{max}}^{\mathrm{obs}} = 2.4805. \tag{3.5}$$

Although this confidence interval is an asymptotic one, as $|\mathbf{X}_0| \to \infty$, it is a very good approximation of the true confidence interval for a finite $|\mathbf{X}_0|$, since $|\mathbf{X}_0|$ is here very large. Since the estimation of θ_0 relies on the values given to the other model parameters $\{p_{\mathrm{mat}}, \alpha, \beta\}$, we evaluate in addition its sensitivity to the values of these parameters. For this purpose,
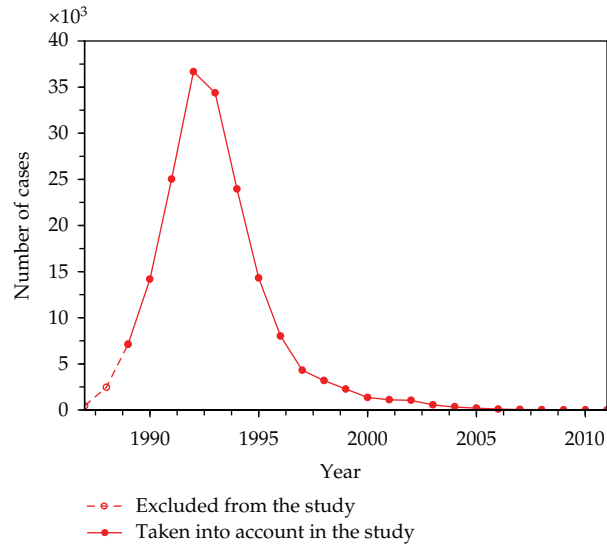
Figure 1: Yearly number of cases of BSE reported in Great Britain from 1987 to 2011 [5]. Our study only takes into account the data from 1989, after the 1988 feed ban law.

Table 1: Sensitivity analysis. Estimation of the infection parameter θ_0, and its confidence interval $[\hat{\theta}_{\min}^{obs}, \hat{\theta}_{\max}^{obs}]$ with asymptotic probability 95%, for different values of the maternal infection parameter p_{mat} and of the latency parameters (α, β). The values $(3.84, 7.46)$ correspond to the Bayesian MAP estimations $(\hat{\alpha}^{MAP}, \hat{\beta}^{MAP})$, and 0.1 to the largest commonly accepted order of magnitude for p_{mat}. The estimations of θ_0 are based on the observed data over the years 1989–2011.

p_{mat}	α	β	$\hat{\theta}_{|\mathbf{x}_0|}^{obs}$	$[\hat{\theta}_{\min}^{obs}, \hat{\theta}_{\max}^{obs}]$
0.1	3.84	7.46	2.4324	$[2.3842, 2.4805]$
0	3.84	7.46	2.4860	$[2.4379, 2.5342]$
1	3.84	7.46	1.9492	$[1.9014, 1.9970]$
0.1	2	7.46	2.7835	$[2.7287, 2.8382]$
0.1	20	7.46	4.0186	$[3.9395, 4.0977]$
0.1	3.84	1	1.0127	$[0.9925, 1.0329]$
0.1	3.84	10	6.2128	$[6.0914, 6.3341]$
0.1	3	6	1.5402	$[1.5095, 1.5710]$
0.1	4	5	1.0227	$[1.0020, 1.0434]$

we compute the estimation of θ_0 and the associated confidence interval with asymptotic probability 95%, for different values of (p_{mat}, α, β) (Table 1). The first line of the table corresponds to the parameters chosen for the model. In each of the four following lines, we fix two coordinates and choose an extremal (unrealistic) value for the third one. It appears that the estimation of θ_0 is almost independent of the value of the maternal infection parameter. However, the estimation seems more strongly dependent on the parameters of the latent period distribution. Nevertheless, even for very unrealistic values (α, β), all the estimations of θ_0 remain in the same order of magnitude of several units. This is really small compared to estimations obtained for the infection *via* Meat and Bone Meal or lactoreplacers (before 1989) which are of the order of 1000 [22]. However, although these estimations are all very small,

θ_0 seems nonnull. This could suggest the existence of a minor but nonnull infection source which is not of maternal type.

3.2.4. Extinction of the Epidemic

We know thanks to Proposition 2.1 that $\{X_n\}_n$ becomes extinct almost surely if and only if $R_0 = \sum_{k=1}^{d} \Psi_k(\theta_0) \leqslant 1$. The estimated basic reproduction is here $R_0(\widehat{\theta}_{|\mathbf{X}_0|}^{\mathrm{obs}}) = 0.1072$. Moreover, solving with a computing program the equation $\sum_{k=1}^{d} \Psi_k(\widehat{\theta}_{|\mathbf{X}_0|}^{\mathrm{obs}})\rho^{-k} = 1$, we obtain the following value for the Perron's root $\rho(\widehat{\theta}_{|\mathbf{X}_0|}^{\mathrm{obs}}) = 0.6665$, which provides the speed of decay of the expected yearly incidence of cases (see (2.6)); from a certain time, the expected number of new cases will decrease from around 33% every year.

3.2.5. Prediction of the Incidences of Cases and Incidences of Infected Cattle

Let us predict the spread of the disease from 2012 by means of simulations of $\{X_n\}_n$, where θ_0 is replaced by its previous estimation $\widehat{\theta}_{|\mathbf{X}_0|}^{\mathrm{obs}} = 2.4324$, and where the initial time of the model is 2011, that is, $\mathbf{X}_0 = \mathbf{X}_{2011}^{\mathrm{obs}}$. The simulations are done recursively using the transition law (3.1). We point out that the model initialized by $\mathbf{X}_0 = \mathbf{X}_{1997}^{\mathrm{obs}}$ provides quite realistic simulations on the period 1998–2011 compared to the real observations on the same period, as illustrated in Figure 2(a). The epidemic process $\{X_n\}_n$ thus seems to provide a satisfying prediction of the overall evolution of the real epidemic. In order to predict the incidences of futures cases, we simulate 1000 trajectories of $\{X_n\}_n$ initialized by the observed values $\mathbf{X}_{2011}^{\mathrm{obs}}$, with the estimated infection parameter $\widehat{\theta}_{|\mathbf{X}_0|}^{\mathrm{obs}} = 2.4324$. We illustrate in Figure 2(c), for each year from 2012, the maximum, minimum, median, 2.5%, and 97.5% quantiles associated with these 1000 realizations. It is also relevant to study and predict the evolution of the incidence of infected cattle in the population, which represents the hidden face of the epidemic. The incidence Z_n of infected cattle at time n, conditionally on the number X_n of cases at that time, is given by the Poisson distribution (3.2). For every $n \geqslant 2012$ and for each of the 1000 previously simulated values X_n, we generate one realization of Z_n. We then illustrate in Figure 2(d), the yearly maximum, minimum, median, 2.5%, and 97.5% quantiles associated with the 1000 realizations.

3.2.6. Prediction of the Year of Extinction

Let $T := 2011 + \inf\{n \geqslant 1, X_n = 0\}$ denote the extinction year of the epidemic process $\{X_n\}_n$. According to (2.7) and denoting by $\mathbf{f}_{\theta_0}$ the offspring generating function of $\{X_n\}_n$ defined in (2.2) (from now on we let the dependence in θ_0 appear in the notation), we have $\mathbb{P}_{\theta_0}(T \leqslant 2011 + n) = \mathbf{f}_{\theta_0,n}(0)^{\mathbf{X}_{2011}^{\mathrm{obs}}}$, for every $n \geqslant 1$, which by iterating $\mathbf{f}_{\theta_0}$ can be computed explicitly. We obtain in particular, for the estimated value $\widehat{\theta}_{|\mathbf{X}_0|}^{\mathrm{obs}} = 2.4324$, the following p-quantiles for the extinction time (see (2.8)): $n_{0.5}^{T} = 2028$, $n_{0.95}^{T} = 2035$ and $n_{0.99}^{T} = 2039$. Keeping in mind that T corresponds to the complete extinction of the epidemic (i.e., $d = 9$ consecutive years without any case), these results actually mean that for the infection parameter $\theta_0 = 2.4324$, with probability larger than 50% (resp., 95% and 99%), no case will arise in the population from year 2020 (resp., 2027 and 2031). Moreover, in order to take into account the uncertainty around the estimation $\widehat{\theta}_{|\mathbf{X}_0|}^{\mathrm{obs}}$ of the infection parameter θ_0, we make use of the asymptotic
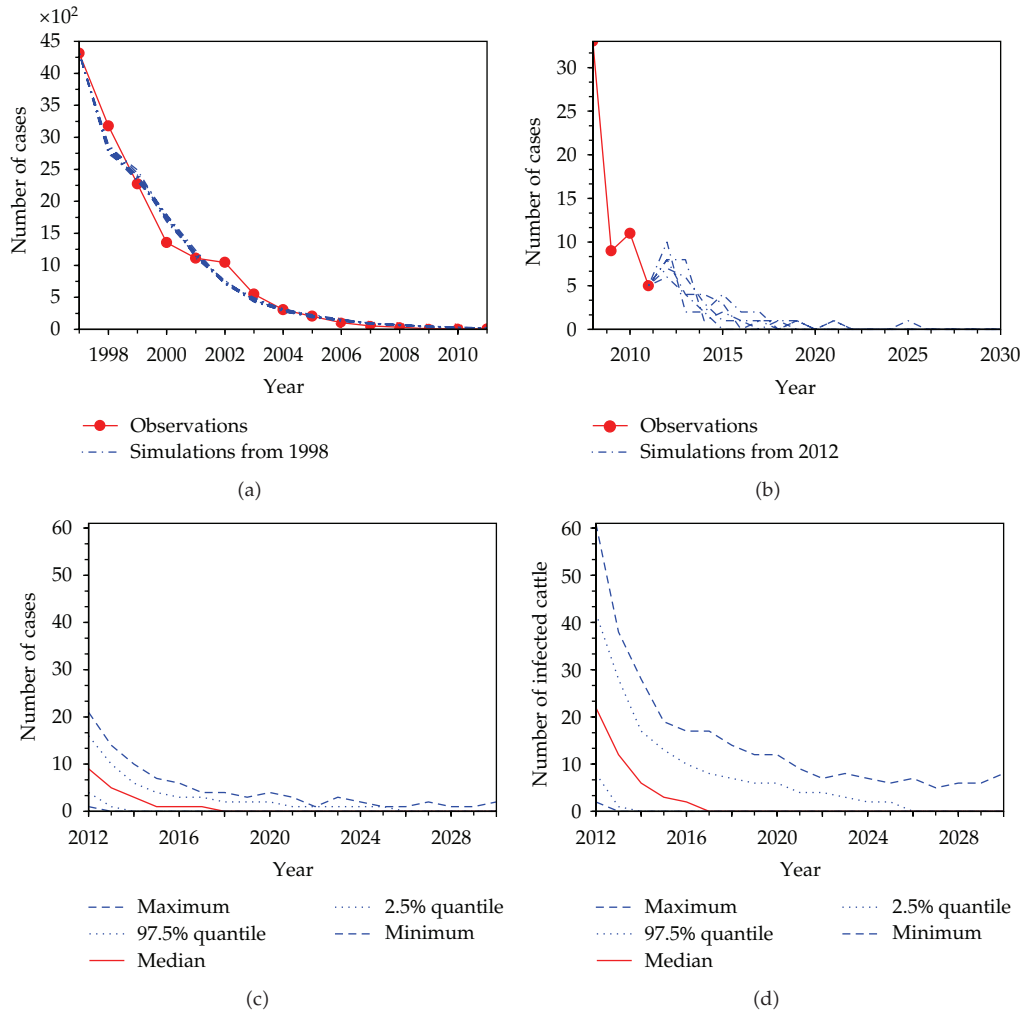
Figure 2: Figure 2(a) 10 simulations of $\{X_n\}_n$ initialized by $\mathbf{X}_{1997}^{\mathrm{obs}}$, and comparison with the observations on the period 1998–2011. Figure 2(b) 5 simulations of $\{X_n\}_n$ initialized by $\mathbf{X}_{2011}^{\mathrm{obs}}$. Figures 2(c) and 2(d) prediction, based on 1000 simulations of the process of the yearly incidences of cases (resp., infected cattle) from 2012. 95% of the trajectories remain in the band delimited by the blue dotted lines. All the simulations are done with the infection parameter $\widehat{\theta}_{|\mathbf{X}_0|}^{\mathrm{obs}} = 2.4324$.

confidence interval (3.5) of θ_0 and of the fact that $\theta \mapsto \mathbb{P}_\theta(T \leqslant n)$ is a decreasing function of θ, which implies that for every $n \geqslant 2011$,

$$\mathbb{P}\left(\mathbb{P}_{\theta_0}(T \leqslant n) \in \left[\mathbb{P}_{\widehat{\theta}_{\max}}(T \leqslant n), \mathbb{P}_{\widehat{\theta}_{\min}}(T \leqslant \mathrm{n})\right]\right) \simeq 95\%. \tag{3.6}$$

We collect in Table 2 the observed interval $[\mathbb{P}_{\widehat{\theta}_{\max}^{\mathrm{obs}}}(T \leqslant n), \mathbb{P}_{\widehat{\theta}_{\min}^{\mathrm{obs}}}(T \leqslant n)]$, for each $n \geqslant 2020$ (if $n < 2019$ we have, conditionally on the initial value $\mathbf{X}_{2011}^{\mathrm{obs}}$, $\mathbb{P}(\mathbf{X}_n = \mathbf{0}) = 0$ because of the memory which is not equal to 0). Note that these intervals are very narrow, leading to an accurate estimation of $\mathbb{P}_{\theta_0}(T \leqslant n)$.

Table 2: Cumulative distribution function of the year of extinction computed with the infection parameters $\widehat{\theta}_{\min}^{\mathrm{obs}} = 2.3842$ and $\widehat{\theta}_{\max}^{\mathrm{obs}} = 2.4805$ defined by (3.5). The values in bold character correspond to the p-quantiles n_p^T for $p = 0.5, 0.95$, and 0.99, and to the asymptotic confidence intervals of $\mathbb{P}_{\theta_0}(T \leqslant n_p^T)$ based on (3.6).

n	$\mathbb{P}_{\widehat{\theta}_{\max}^{\mathrm{obs}}}(T \leqslant n)$	$\mathbb{P}_{\widehat{\theta}_{\min}^{\mathrm{obs}}}(\cdots)$	n	$\mathbb{P}_{\widehat{\theta}_{\max}^{\mathrm{obs}}}(\cdots)$	$\mathbb{P}_{\widehat{\theta}_{\min}^{\mathrm{obs}}}(\cdots)$	n	$\mathbb{P}_{\widehat{\theta}_{\max}^{\mathrm{obs}}}(\cdots)$	$\mathbb{P}_{\widehat{\theta}_{\min}^{\mathrm{obs}}}(\cdots)$
2020	0.0000	0.0000	2030	0.7610	0.7807	2040	0.9950	0.9957
2021	0.0000	0.0000	2031	0.8310	0.8465	2041	0.9967	0.9972
2022	0.0010	0.0014	2032	0.8816	0.8934	2042	0.9978	0.9981
2023	0.0121	0.0152	2033	0.9186	0.9273	2043	0.9985	0.9988
2024	0.0496	0.0592	2034	0.9451	0.9513	2044	0.9990	0.9992
2025	0.1211	0.1390	**2035**	**0.9633**	**0.9677**	2045	0.9995	0.9994
2026	0.2325	0.2579	2036	0.9755	0.9786	2046	0.9996	0.9996
2027	0.3756	0.4047	2037	0.9835	0.9857	2047	0.9997	0.9998
2028	**0.5303**	**0.5584**	2038	0.9889	0.9904	2048	0.9998	0.9998
2029	0.6619	0.6862	**2039**	**0.9925**	**0.9936**	2049	0.9999	0.9999

Table 3: Cumulative distribution function of the total size of the epidemic for the infection parameters $\widehat{\theta}_{\min}^{\mathrm{obs}} = 2.3842$ and $\widehat{\theta}_{\max}^{\mathrm{obs}} = 2.4805$ defined by (3.5). The values in bold character correspond to the p-quantiles n_p^N for $p = 0.5, 0.95$, and 0.99, and to the asymptotic confidence intervals of $\mathbb{P}_{\theta_0}(N \leqslant n_p^N)$.

n	$\mathbb{P}_{\widehat{\theta}_{\max}^{\mathrm{obs}}}(N \leqslant n)$	$\mathbb{P}_{\widehat{\theta}_{\min}^{\mathrm{obs}}}(N \leqslant n)$	n	$\mathbb{P}_{\widehat{\theta}_{\max}^{\mathrm{obs}}}(\cdots)$	$\mathbb{P}_{\widehat{\theta}_{\min}^{\mathrm{obs}}}(\cdots)$	n	$\mathbb{P}_{\widehat{\theta}_{\max}^{\mathrm{obs}}}(\cdots)$	$\mathbb{P}_{\widehat{\theta}_{\min}^{\mathrm{obs}}}(\cdots)$
$\cdots$	$\cdots$	$\cdots$	16	0.1293	0.1669	28	0.8702	0.9048
5	0.0001	0.0000	17	0.1795	0.2263	29	0.8942	0.9302
6	0.0002	0.0003	18	0.2373	0.2915	30	0.9283	0.9495
7	0.0005	0.0008	19	0.3025	0.3637	**31**	**0.9480**	**0.9646**
8	0.0013	0.0020	20	0.3741	0.4409	32	0.9635	0.9759
9	0.0027	0.0046	21	0.4486	0.5189	33	0.9742	0.9838
10	0.0060	0.0094	**22**	**0.5231**	**0.5934**	34	0.9824	0.9892
11	0.0119	0.0178	23	0.5975	0.6639	**35**	**0.9881**	**0.9928**
12	0.0214	0.0310	24	0.6655	0.7287	36	0.9917	0.9952
13	0.0365	0.0518	25	0.7282	0.7840	37	0.9945	0.9968
14	0.0585	0.0801	26	0.7829	0.8323	38	0.9964	0.9980
15	0.0891	0.1181	27	0.8304	0.8726	39	0.9977	0.9989

3.2.7. Prediction of the Epidemic Size

Let $N := \sum_{n=1}^{T-2011} X_n$ be the total size of the future epidemic from 2012 (total number of cases from 2012 until the extinction of the epidemic). We compute the distribution of N using (2.11), conditionally on the event $\{\mathbf{X}_0 = \mathbf{X}_{2011}^{\mathrm{obs}}\}$. We obtain in particular, for the estimated value $\widehat{\theta}_{|\mathbf{X}_0|}^{\mathrm{obs}} = 2.4324$, the following p-quantiles for the total epidemic size (see (2.12)): $n_{0.5}^N = 22$, $n_{0.95}^N = 31$ and $n_{0.99}^N = 35$. From (2.13) we deduce that the mean value and variance of N for the parameter $\widehat{\theta}_{\min}^{\mathrm{obs}}$ (resp., $\widehat{\theta}_{\max}^{\mathrm{obs}}$) are 21 and 27 (resp., 22 and 28). Moreover, we obtain similarly as for the extinction time and thanks to (3.5) a confidence interval $[\mathbb{P}_{\widehat{\theta}_{\max}}(N \leqslant n), \mathbb{P}_{\widehat{\theta}_{\min}}(N \leqslant n)]$ of $\mathbb{P}_{\theta_0}(N \leqslant n)$, for every $n \in \mathbb{N}$, with asymptotic probability 95%, collected in Table 3.

3.2.8. Study of the Very Late Extinction Case

In order to predict the behavior of the "most dangerous" evolution of the epidemic, we first need to compute the estimation $\widehat{\theta}_n^{*\mathrm{obs}}$ (2.47) based on the data in Great Britain. The number of available observations is only $n = 14$, so we are far from the asymptotic setting $n \rightarrow \infty$ of Theorem 2.7. However, the large value of $|\mathbf{X}_0|$ can make us hope for a good accuracy. We point out that, by making use of the estimator $\widehat{\theta}_n^*$ on the real data, we make an unverifiable assumption on the future of the epidemic; we consider the data as if they were the beginning of a trajectory with very late extinction. This should have the following consequence: the estimation provided by $\widehat{\theta}_n^{*\mathrm{obs}}$ should *a priori* be a bit smaller than the value 2.4324 provided by $\widehat{\theta}_{|\mathbf{X}_0|}^{\mathrm{obs}}$. Indeed we obtain $\widehat{\theta}_n^{*\mathrm{obs}} = 2.4305$. Using this value, we deduce from (2.51) the confidence interval $[\widehat{\theta}_{\min}^*, \widehat{\theta}_{\max}^*]$ of θ_0 with asymptotic probability 95%, where $\widehat{\theta}_{\min}^* := \widehat{\theta}_n^* - 1.96\widehat{c}_2^{-1}$, $\widehat{\theta}_{\max}^* := \widehat{\theta}_n^* + 1.96\widehat{c}_2^{-1}$, and

$$\widehat{c}_2 := \sum_{k=0}^n \left(f'\left(\widehat{\theta}_n^*, \mathbf{X}_k^*\right)\right)^2 \left[\sum_{k=0}^n \left(f'\left(\widehat{\theta}_n^*, \mathbf{X}_k^*\right)\right)^2 g\left(\widehat{\theta}_n^*, \mathbf{X}_k^*\right)\right]^{-1/2}. \tag{3.7}$$

Therefore, $\widehat{c}_2^{\mathrm{obs}} = 40.6988$, and

$$\mathbb{P}\left(\theta_0 \in \left[\widehat{\theta}_{\min}^*, \widehat{\theta}_{\max}^*\right]\right) \simeq 95\%, \qquad \widehat{\theta}_{\min}^{*\mathrm{obs}} = 2.3823, \qquad \widehat{\theta}_{\max}^{*\mathrm{obs}} = 2.4787, \tag{3.8}$$

which is of the same magnitude order as the confidence interval $[2.3842, 2.4805]$ obtained with the unconditioned process (see (3.5)). Let us predict the most dangerous evolution thanks to the transition law of the conditioned process given by Proposition 2.2. Note that if one computes the eigenvalues ρ and λ introduced in Proposition 2.5, one obtains $\rho(\widehat{\theta}^{*\mathrm{obs}}) = 0.6664$ and $|\lambda(\widehat{\theta}^{*\mathrm{obs}})| = 0.5570$. It thus appears that the convergence of the epidemic process conditioned on nonextinction at time k to the Q-process as $k \rightarrow \infty$ is not very fast. As a consequence the study of the Q-process only provides information about the behavior of the disease spread in the case of an extremely late extinction. First, we see thanks to Figure 3(a) that the simulations provided by the conditioned process initialized by $\mathbf{X}_0 = \mathbf{X}_{1997}^{\mathrm{obs}}$, and where θ_0 is estimated by $\widehat{\theta}_n^{*\mathrm{obs}} = 2.4305$, are true to the real observations on the period 1999–2011. Figure 3(b) is an example of one simulation in the period 2012–2040 of the conditioned process, for $\mathbf{X}_0^* = \mathbf{X}_{2011}^{\mathrm{obs}}$. It appears that the values of this simulated trajectory are rapidly very small and of course are never equal to 0 for $d = 9$ consecutive times. For a finer prediction, we simulate 1000 realizations of this process from 2012 until 2050, with $\mathbf{X}_0^* = \mathbf{X}_{2011}^{\mathrm{obs}}$. Moreover, for every $n \geqslant 2012$ and for each of the 1000 simulated values X_n^*, we make one realization of the incidence Z_n of infected cattle at time n, according to the law given by (3.2). Figures 3(c) and 3(d) represent the yearly maximum, minimum, median, 2.5%, and 97.5% quantiles associated with the 1000 realizations of, respectively, the incidence of cases and infected cattle, in case of an extreme late extinction. It appears thanks to Figures 3(c) and 3(d) that the supposedly "most dangerous" trajectories nevertheless do not reach high values and do not present a new peak of epidemic.
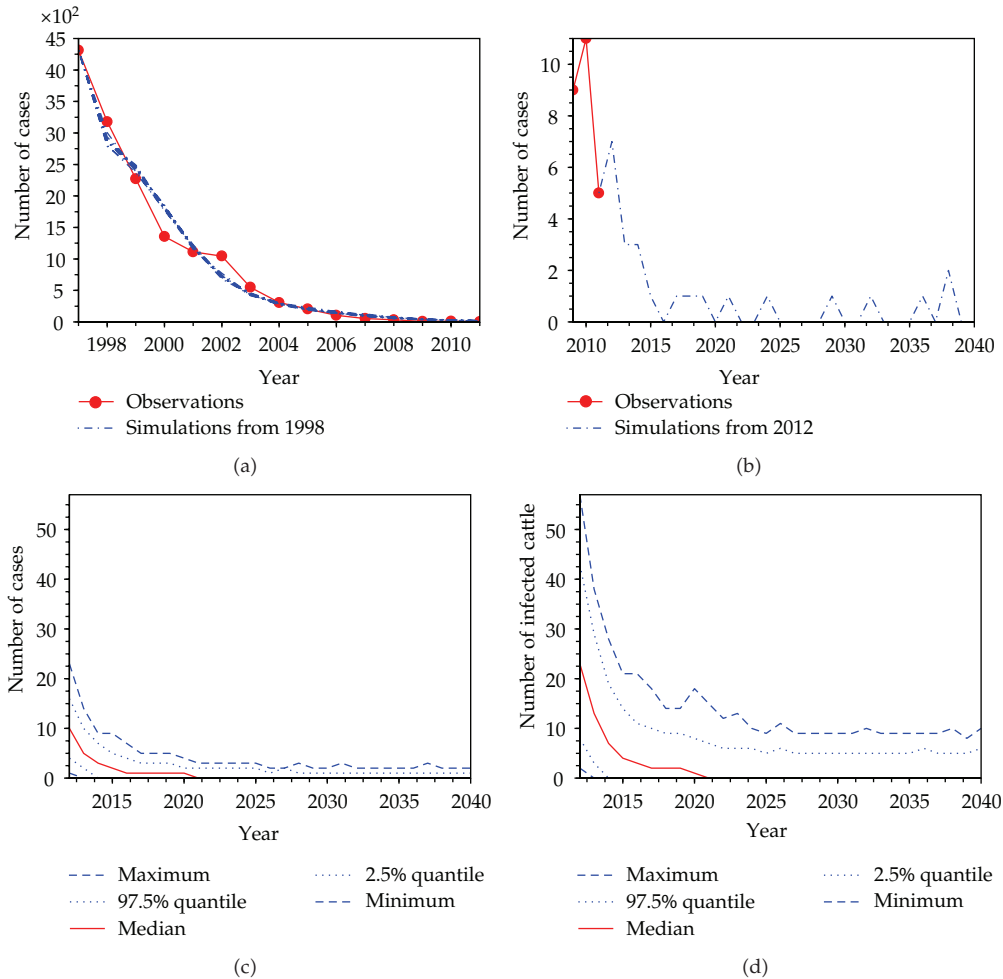
Figure 3: Figure 3(a) 10 simulations of $\{X_n^*\}_n$ initialized by $\mathbf{X}_{1997}^{\text{obs}}$, and comparison with the observations on the period 1998–2011. Figure 3(b) one simulation of $\{X_n^*\}_n$ initialized by $\mathbf{X}_{2011}^{\text{obs}}$. Figures 3(c) and 3(d) prediction, based on 1000 simulations of the conditioned process of the yearly incidences of cases (resp., infected cattle) from 2012. 95% of the trajectories remain in the band delimited by the blue dotted lines. All the simulations are done with the infection parameter $\widehat{\theta}_{|\mathbf{X}_0|}^{*\text{obs}} = 2.4305$.

4. Conclusion

The methodology developed in Section 2 thus applies well here and enables not only to confirm mathematically what is commonly accepted, namely, that BSE is fading out, but also to predict, with a very large probability, that the last BSE case will occur before 2027, and that until the complete extinction of the epidemic in the population, there will be less than 31 cases to come. We obtain moreover from Figure 2(d) the order of magnitude of the number of infected cattle in the population. In addition, the estimation of the infection parameter concludes to the possible existence of a minor but nonnull infection source which is not of maternal type, and which is very small (only around 3 newly infected animal per year and per infective) compared to the main source of horizontal infection until 1988 due to protein

supplements. Finally, the study of the worst-case scenario shows that even in the case of an extreme late extinction of the disease in the population, the incidence of cases will decrease quite rapidly to 0, with afterwards only 1 or 2 yearly cases occurring regularly but sparsely, with no appearance of a new peak of epidemic. We have shown with this example that the methodology developed in Section 2 provides accurate tools to study the decay phase of an epidemic under the current sanitary measures, which would help to make new policy decisions. This evaluation is all the more relevant since it is obtained not by simply computing what should most probably happen, but also by taking into account the variability of many factors (infection, incubation, and survival), and by studying the potentially most dangerous evolution.

References

[1] V. M. Panaretos, "Partially observed branching processes for stochastic epidemics," *Journal of Mathematical Biology*, vol. 54, no. 5, pp. 645–668, 2007.

[2] P. Trapman, R. Meester, and H. Heesterbeek, "A branching model for the spread of infectious animal diseases in varying environments," *Journal of Mathematical Biology*, vol. 49, no. 6, pp. 553–576, 2004.

[3] F. Ball, "The threshold behaviour of epidemic models," *Journal of Applied Probability*, vol. 20, no. 2, pp. 227–241, 1983.

[4] F. Ball and P. Donnelly, "Strong approximations for epidemic models," *Stochastic Processes and their Applications*, vol. 55, no. 1, pp. 1–21, 1995.

[5] OIE (World Organisation for Animal Health), "Number of cases of bovine spongiform encephalopathy (BSE) reported in the United Kingdom," 2012, http://www.oie.int/en/animal-health-in-the-world/bse-specific-data/.

[6] K. B. Athreya and P. E. Ney, *Branching Processes*, Springer, New York, NY, USA, 1972.

[7] P. Haccou, P. Jagers, and V. A. Vatutin, *Branching Processes: Variation, Growth, and Extinction of Populations*, Cambridge Studies in Adaptive Dynamics, Cambridge University Press, Cambridge, UK, 2007.

[8] A. Joffe and F. Spitzer, "On multitype branching processes with $\rho \leq 1$," *Journal of Mathematical Analysis and Applications*, vol. 19, pp. 409–430, 1967.

[9] C. Jacob and S. Pénisson, "Limit models for branching processes with memory and population dependence in large populations," Technical Report UR341, INRA Jouy-en-Josas, 2012.

[10] S. Dallaporta and A. Joffe, "The Q-process in a multitype branching processes," *International Journal of Pure and Applied Mathematics*, vol. 42, no. 2, pp. 235–240, 2008.

[11] J. N. Darroch and E. Seneta, "On quasi-stationary distributions in absorbing discrete-time finite Markov chains," *Journal of Applied Probability*, vol. 2, pp. 88–100, 1965.

[12] T. E. Harris, "Branching processes," *Annals of Mathematical Statistics*, vol. 19, pp. 474–494, 1948.

[13] V. P. Godambe and C. C. Heyde, "Quasi-likelihood and optimal estimation," *International Statistical Review*, vol. 55, no. 3, pp. 231–244, 1987.

[14] J. R. Blum, D. L. Hanson, and J. I. Rosenblatt, "On the central limit theorem for the sum of a random number of independent random variables," *Probability Theory and Related Fields*, vol. 1, pp. 389–393, 1962-1963.

[15] M. P. Quine and P. Durham, "Estimation for multitype branching processes," *Journal of Applied Probability*, vol. 14, no. 4, pp. 829–835, 1977.

[16] P. Billingsley, *Statistical Inference for Markov Processes*, Statistical Research Monographs, Vol. II, The University of Chicago Press, Chicago, Ill, USA, 1961.

[17] C. Jacob, "Conditional least squares estimation in nonstationary nonlinear stochastic regression models," *The Annals of Statistics*, vol. 38, no. 1, pp. 566–597, 2010.

[18] R. Rebolledo, "Central limit theorems for local martingales," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 51, no. 3, pp. 269–286, 1980.

[19] B. L. S. Prakasa Rao, *Semimartingales and Their Statistical Inference*, vol. 83 of *Monographs on Statistics and Applied Probability*, Chapman & Hall/CRC, 1999.

[20] G. A. Wells, A. C. Scott, C. T. Johnson et al., "A novel progressive spongiform encephalopathy in cattle," *Veterinary Record*, vol. 121, no. 18, pp. 419–420, 1987.

[21] M. Kaatz, C. Fast, U. Ziegler et al., "Spread of classic BSE prions from the gut via the peripheral nervous system to the brain," *The American Journal of Pathology*, vol. 181, no. 2, pp. 515–524, 2012.

[22] C. Jacob, L. Maillard-Teyssier, J.-B. Denis, and C. Bidot, "A branching process approach for the propagation of the Bovine Spongiform Encephalopathy in Great-Britain," in *Workshop on Branching Processes and their Applications*, vol. 197 of *Lecture Notes in Statistics*, pp. 225–240, Springer, Berlin, Germany, 2010.

[23] J. Bradbury, "Maternal transmission of BSE demonstrated in cattle," *The Lancet*, vol. 348, no. 9024, p. 393, 1996.

[24] C. A. Donnelly, N. M. Ferguson, A. C. Ghani, M. E. J. Woolhouse, C. J. Watt, and R. M. Anderson, "The epidemiology of BSE in cattle herds in Great Britain. I. Epidemiological processes, demography of cattle and approaches to control by culling," *Philosophical Transactions of the Royal Society B*, vol. 352, no. 1355, pp. 781–801, 1997.

[25] C. A. Donnelly, N. M. Ferguson, A. C. Ghani, J. W. Wilesmith, and R. M. Anderson, "Analysis of dam-calf pairs of BSE cases: confirmation of a maternal risk enhancement," *Proceedings of the Royal Society B*, vol. 264, no. 1388, pp. 1647–1656, 1997.