Spectral method in epidemic time series

Jacques Demongeot[(a)] and Pierre Magal[(b),*]
[(a)] *Université Grenoble Alpes, AGEIS EA7407, F-38700 La Tronche, France*
[(b)] *Univ. Bordeaux, IMB, UMR 5251, F-33400 Talence, France.*
CNRS, IMB, UMR 5251, F-33400 Talence, France.

September 11, 2022

Abstract

We start this article by deriving an autoregressive moving average model from a discrete-time epidemic model involving the age of infection. The deterministic part of such a model is nothing but a linear scalar delay difference equation. The article's main idea is to use the spectrum (or part of the spectrum) associated with this difference equation's characteristic equation to describe the data and the model. Next, we present some results of identification of the model's parameters when all the eigenvalues are known. We apply these results to the exponential growth phase for Japan's third epidemic wave of COVID-19. We start by considering one week and extend our analysis to one month. We identify the several shapes for daily reproduction numbers in both cases using only a few eigenvalues to fit the data.

Keywords: *Epidemic models, Time series, Spectral method, Spectral truncation method, Phenomenological models*

1 Introduction

In the present paper, we reconsider a day-by-day discrete time epidemic model with age of infection presented in Demongeot et al. [9]. This model is a discrete time version of the Volterra integral formulation of the Kermack-McKendrick model with age of infection [21]. The variation of the number of susceptible individuals $S(t)$ is given each day $t = t_0, t_0 + 1, \ldots$, by

$$S(t) = S_0 - \sum_{d=t_0}^{t-1} N(d), \forall t \geq t_0, \tag{1.1}$$

where $S(t)$ is the number of susceptible individuals at time t, and $N(t)$ is the daily number of new infected at time t. Throughout the paper, we use the following convention for the sum

$$\sum_{d=k}^{m} = 0, \text{ whenever } m < k.$$

As a consequence, when $t = t_0$ equation (1.1) gives

$$S(t_0) = S_0.$$

We assume for simplicity that the epidemic starts from a single cohort of infected at time t_0, then the number of infectious individuals is given by

$$I(t) = \left[\Gamma(t - t_0)I_0 + \sum_{d=1}^{t-t_0} \Gamma(d) \times N(t - d)\right], \tag{1.2}$$

*Corresponding author. e-mail: pierre.magal@u-bordeaux.fr

where I_0 is the number of infected individuals at time t_0, and $\Gamma(d)$ is the probability for an infected to be infectious after d day of infection.

We assume that $N(t)$ the number of new infected at time t is the product of the transmission rate $\tau(t)$ with $S(t)$ the number susceptible individuals and $I(t)$ the number of infectious at time t. That is

$$N(t) = \tau(t)\, S(t)\, I(t). \tag{1.3}$$

By replacing $I(t)$ by the right hand side of (1.2) in (1.3), we obtain

$$N(t) = \tau(t)S(t)\left[\Gamma(t-t_0)I_0 + \sum_{d=1}^{t-t_0} \Gamma(d) \times N(t-d)\right]. \tag{1.4}$$

Now assuming that $t \to \tau(t) = \tau_0$ and $t \to S(t) = S_0$ are constant (over a short period of time), then we define the **daily reproduction numbers** as

$$R_0(d) = \tau_0\, S_0\, \Gamma(d), \forall d \geq 0.$$

The quantity $R_0(d)$ is the average number of secondary infected produced by a single infected on the day d since infection (see [9] for more details). Therefore, the **basic reproduction number** is the following quantity

$$\mathcal{R}_0 = \sum_{d=1}^{n} R_0(d), \tag{1.5}$$

where n is the maximal duration of the infection.

Moreover when $t \to \tau(t) = \tau_0$ and $t \to S(t) = S_0$ are constant, the equation (1.4) becomes a linear discrete time Volterra integral equation

$$N(t) = \underbrace{R_0(t-t_0) \times I_0}_{\text{(I)}} + \underbrace{\sum_{d=1}^{t-t_0} R_0(d) \times N(t-d), \forall t \geq t_0,}_{\text{(II)}} \tag{1.6}$$

where (I) is the number of infected produced directly by the I_0 infected individuals already present on day t_0, and (II) is the number of new infected individuals at time t produced by the new infected individuals since day t_0.

If we consider the first terms of the discrete time Volterra equation (1.6), we obtain

$$\begin{aligned}
N(t_0) \quad &= R_0(0) \times I_0,\\
N(t_0+1) \quad &= R_0(1) \times I_0 + R_0(1) \times N(t_0),\\
N(t_0+2) \quad &= R_0(2) \times I_0 + R_0(2) \times N(t_0) + R_0(1) \times N(t_0+1),\\
N(t_0+3) \quad &= R_0(3) \times I_0 + R_0(3) \times N(t_0) + R_0(2) \times N(t_0+1) + R_0(1) \times N_0(t_0+2),\\
&\vdots
\end{aligned}$$

In practice, we can assume that $R_0(0) = 0$ since infected individuals are not infectious immediately after being infected. Under this additional assumption, we obtain the system

$$\begin{aligned}
N(t_0) \quad &= 0,\\
N(t_0+1) \quad &= R_0(1) \times I_0,\\
N(t_0+2) \quad &= R_0(2) \times I_0 + R_0(1) \times N(t_0+1),\\
N(t_0+3) \quad &= R_0(3) \times I_0 + R_0(2) \times N(t_0+1) + R_0(1) \times N_0(t_0+2),\\
&\vdots
\end{aligned}$$

Therefore, can be rewritten a scalar delay difference equation

$$N(t) = R_0(1)N_0(t-1) + \ldots + R_0(t-t_0-1)N_0(t-(t-t_0-1)) + R_0(t-t_0)I_0, \forall t \geq t_0. \tag{1.7}$$

Assume that the infectious period has a finite length n. That is

$$R_0(a) = 0, \forall a \geq n+1.$$

Then by defining $t_1 = t_0 + n + 1$, the equation (1.6) becomes

$$N(t) = \sum_{d=1}^{n} R_0(d) \times N(t-d), \forall t \geq t_1, \tag{1.8}$$

with the initial values

$$N(t) = N_0(t), \forall t \in [t_1 - n, t_1]. \tag{1.9}$$

The goal of this article is to understand how to identify the daily reproduction numbers $d \in \{1, \ldots, n\} \to R_0(d)$ in (1.8) knowing $t \in [t_1, t_2] \to N(t)$ on some finite time interval. This problem is particularly important to derive the average dynamic of infection at the level of a single patient.

The literature about parameters identification for epidemic model with age of infection can be divided in two groups depending on the assumptions made. The first group assumes that $d \to \Gamma(d)$ is a given function and they estimate the time dependent transmission rate $t \to \tau(t)$. As a consequence, we obtain the instantaneous (daily or effective) reproduction number, which is

$$\mathcal{R}_0(t) = \tau(t)S(t)\sum_{d=1}^{n}\Gamma(d).$$

We refer to [1], [2], [3], [8], [14], [15], [29], [30] (and references therein) for more results about this subject.

The second group corresponds to the assumptions considered here. That is we assume that $t \to \tau(t) = \tau_0$ and $t \to S(t) = S_0$ constant functions (over a short period of time) and estimate the daily reproduction number. That is the case for the discrete time model in [41] and more recently for the continuous time model in [9]. The major default in [41] is that the estimated $d \to R_0(d)$ does not remain positive. We will have the same problem is Section 3.1 when we will the full spectrum. In Section 3.2, to solve this problem, we introduce a method using the dominant and secondary eigenvalue only.

This article aims to investigate the shape of the distribution $d \to R_0(d)$ from the data of COVID-19. In Figure 1 we illustrate the notion of U or M shape distribution.

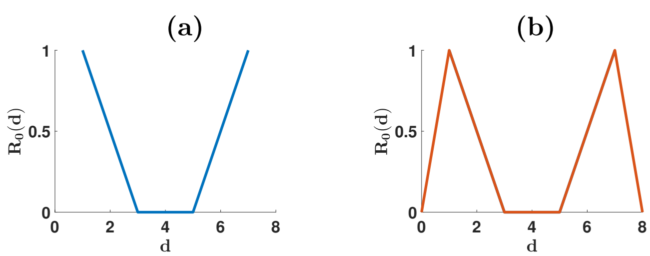

Figure 1: *In this figure, we illustrate the notion of U shape distribution in (a) and M shape distribution in (b). Recall that $R_0(d)$ represents the ability of patients to transmit the pathogen after d days since they got infected. The U shape or M shape distribution means that patients can transmit the pathogen since the beginning of their infection. Then they become less infectious in the middle of the infected period. Finally, they become infectious again at the end of the infected period. The only difference between U and M shape distribution is to include days* 0 *and* 8 *and* $R_0(0) = R_0(8) = 0$ *in the plot.*

The U or M shape distribution are well known in the context of influenza [6] [19]. In Figure 3, we present some figures reflecting patients' viral load for COVID-19.

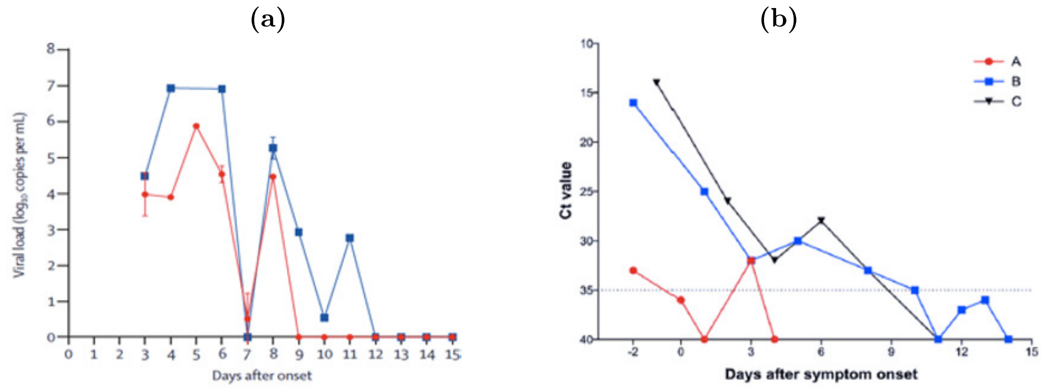

Figure 2: *Viral load in COVID-19 real patients [31]. In figure (a) the red curve corresponds to the throat swab and the blue curve corresponds to the sputum. In figure (b) the curves correspond to several patients (A), (B), and (C).*

Such U shape has not yet systematically studied in COVID-19 data, but observations of the evolution of the viral load have been done in some patients and show this U-shape. The Figure 3 shows such a U shaped evolution for the viral load in real [31].

The present work is directly connected to the original work of Peter Whittle in 1951 [46] [47] who introduced the Auto Regressive Moving Average (ARMA) model, after the seminal paper on time series by N. Wiener [42],

$$N(t) = \underbrace{K(1)N(t-1) + K(2)N(t-2) + \ldots + K(n)N(t-n)}_{\text{Auto regressive part}} + \underbrace{w(t)}_{\text{Moving average part}}, \tag{1.10}$$

where $N(t)$ is the size at time t of the population whose growth is forecasted, the kernel $d \to K(d)$ has real values, n is the regression order, and here $w(t)$ stands for a noise. The equation (1.10) has been extensively studied under the denomination of ARMA models by many authors [5], [26], [33], [34], [35], [38] [39].

Here, we propose a new approach based on the spectral properties of the population growth equation to capture information from data. Our goal is to estimate the shape of the daily reproduction numbers $d \to R_0(d)$. Spectral methods are not new. But it usually refers to Fourier transform the with frequencies associated to various period see Priestley [32] [33]. Spectral methods are not new. But it usually refer to Fourier transform with frequencies associated to various periods, corresponding to a fundamental period and its sub-multiples (harmonics). If we consider the auto regressive part only, the spectrum of the delay difference equation is determined by its characteristic equation

$$\lambda^n = K(1)\lambda^{n-1} + K(2)\lambda^{n-2} + \ldots + K(n-1)\lambda + K(n).$$

The main idea in this article is to use these eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{C}$ (i.e. the solution of the characteristic equation) to identify the parameters $K(1), K(2), \ldots, K(n)$. The eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{C}$ are estimated by some separated method. In section 2 we will see that when all the eigenvalues are non null and separated two by two then we can compute the parameters $K(1), K(2), \ldots, K(n)$ by using the eigenvalue only.

The idea of using eigenvalues in population dynamics goes back to Malthus [28], who, in 1798, first identified in a mixture of populations, the one that would impose itself on the others, because having the exponential growth of exponent the largest, this leading exponent having been called Malthusian parameter by Fisher [13]. The Malthusian growth seeming unrealistic, the saturation logistic term, was introduced further by Lambert [24] and then, extending the initial work by Euler [12], Lotka [27], Leslie [25] and Hahn [17] gave the current matrix form of the discrete population growth equations.

But as far as we know estimating the subdominant eigenvalues to characterize the system is new. So the key idea of this work is to use the dominant eigenvalue λ_1 and also the following pair of complex conjugated eigenvalues $\lambda_2, \overline{\lambda}_2$ as an estimator to reconstruct the kernel of the auto regressive part.

This work is motivated by the times series provided by the daily number of reported cases data for COVID-19. During the COVID-19 pandemic, most people viewed the oscillation around the exponential growth at the beginning of an epidemic wave as the default in reporting the data. The error is probably partly due to the reporting data process (random noise). Nevertheless, a significant remaining part of such oscillations could be connected to the infection dynamic at the level of a single average patient. Eventually, the central question we try to address here is: Is there some hidden information in the signal around the exponential tendency for COVID-19 data? So consider the early stage of an epidemic phase, and we try to exploit the oscillations around the tendency in order to reconstruct the infection dynamic at the level of a single average patient.

We start by investigating the connection between a signal decomposed into a sum of damped or amplified oscillations and a renewal equation. The prototype example we have in mind is the following

$$N(t) = A_1 e^{\alpha_1 t} + e^{\alpha_2 t} \left[A_2 \cos(\omega_2 t) + B_2 \sin(\omega_2 t) \right] + C, \; \forall t \geq t_1 - n,$$

where $A_1, A_2, A_3 \in \mathbb{R}$, $\alpha_1 > 0$, $\alpha_2 \in \mathbb{R}$, and $\omega_2 > 0$.

In Figure 3, we illustrate a growing function with damped oscillations (i.e., $\alpha_2 < 0$) and amplified oscillations (i.e., $\alpha_2 > 0$). It is clear from Figure 3 that a periodic function can not represent such a signal, and extending such a signal by periodicity would be artificial. Indeed, the Fourier decomposition would only provide purely imaginary eigenvalues that would exclude a continuation of the exponential growth (i.e., eigenvalues with non-zero real parts). To apply wavelets theory (see, for example, in [4]), we need to extend the data for negative times by symmetry with respect to the initial time $t = 0$, and we need a decreasing function ($\alpha_1 < 0$ and $\alpha_2 < 0$).

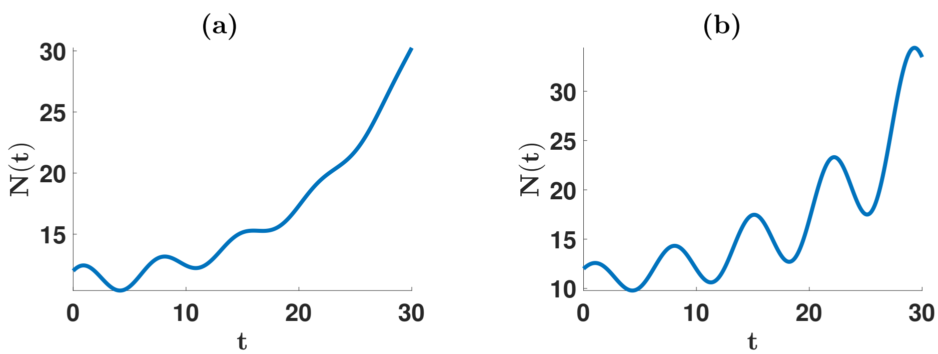

Figure 3: *We plot an exponentially growing function (a) damped oscillations, and (b) amplified oscillations.*

Here, we are more interested in the model resulting from the data (i.e., that is $R_0(d) \geq 0, \forall d = 1, \ldots, n$) than in the fit of the data. The major problem with the Fourier method is that this method provides only eigenvalues with zero real parts (that is due to the periodicity required for this method). Such eigenvalues are well adapted to a periodic signal, but this does not suitable to describe for example an ever-growing function (as in Figure 3). Consequently, the Fourier method is not well adapted to derive a non-negative daily reproduction number (i.e., $R_0(d) \geq 0, \forall d = 1, \ldots, n$).

The plan of the paper is the following. In Section 2, we present some identification results for the daily reproduction numbers. In Section 2.3, we present some phenomenological models that will be compared to the data. In Section 3, we fit the phenomenological models to the cumulative numbers of reported cases in Japan over a period of 10 days.

2 Materials and methods

2.1 Non identifiability result

The Leslie matrix associated to the difference equation (1.8) is

$$L = \begin{pmatrix} R_0(1) & R_0(2) & R_0(3) & \cdots & R_0(n) \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}. \tag{2.1}$$

The **characteristic equation** of (2.1) is

$$\lambda^n = \sum_{d=1}^{n} R_0(d)\lambda^{n-d}, \tag{2.2}$$

for $\lambda \in \mathbb{C}$, which is equivalent to (whenever $\lambda \neq 0$)

$$1 = \sum_{d=1}^{n} R_0(d)\lambda^{-d}.$$

From the above formula, we deduce that (2.2) has exactly one positive eigenvalue. By the Perron-Frobenius theorem applied to the Leslie matrix L, we know that (by considering the norm of linear operator)

$$r(L) := \lim_{n\to+\infty} \|L^n\|_{\mathcal{L}(\mathbb{R})}^{1/n} > 0,$$

the spectral radius of L is the unique positive solution of (2.2). Moreover all the remaining eigenvalues have a modulus smaller or equal to $r(L)$. We refer to [11, Chapter 4] for more results about this subject.

Non identifiability result: Let $\lambda_\star > 0$ and $N_\star \neq 0$. Then

$$N(t) = N_\star \lambda_\star^{t-t_1}, \forall t \geq t_1,$$

is a known solution of (1.8) if and only if $\lambda_\star$ is a solution of the characteristic equation.

Assume that $d \in [1, n] \to R^\star(d) \geq 0$ is given, and satisfies

$$\sum_{d=1}^{n} R^\star(d) > 0.$$

Then if we define

$$R_0(a) = \frac{R^\star(a)}{\sum_{d=1}^{n} R^\star(d)\lambda_\star^{-d}}, \forall a = 1, \ldots, n,$$

we deduce that the equation (2.2) is satisfied for $\lambda = \lambda_\star$, and $N(t) = N_\star \lambda_\star^{t-t_1}$ is a solution of (1.8). We conclude that a single function $N(t) = N_\star \lambda_\star^{t-t_1}$ is not enough to identify $R_0(1), R_0(2), R_0(3), \ldots, R_0(n)$.

2.2 Identifiability result

Assumption 2.1. *Assume that $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ are nonzero complex numbers, and are separated two by two. That is*

$$\lambda_i \neq 0, \forall i = 1, \ldots, n.$$

and

$$\lambda_i \neq \lambda_j, \text{ whenever } i \neq j.$$

Remark 2.2. *Since the coefficients of the characteristic equation (2.2) are all real, we could also impose that the conjugate of each eigenvalue belongs to the spectrum. That is*

$$\overline{\lambda}_i \in \{\lambda_1, \dots, \lambda_n\}, \forall i = 1, \dots, n.$$

But that is not necessary in this subsection.

Remark 2.3. *When all the eigenvalues are real, the above assumption will be satisfied if and only if $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ are nonzero real numbers which are separated two by two. Up to a permutation, that is*

$$\lambda_i \neq 0, \forall i = 1, \dots, n,$$

and

$$\lambda_1 < \lambda_2 < \dots < \lambda_n.$$

Lemma 2.4. *Let Assumption 2.1 be satisfied. Assume that each λ_i satisfies the characteristic equation (2.2). Then the Leslie matrix L defined by (2.1) is diagonalizable and invertible, for each $U_1, U_2, \dots, U_n \in \mathbb{C}$,*

$$U(t) = U_1 \lambda_1^t + U_2 \lambda_2^t + \dots + U_n \lambda_n^t, \ \forall t \geq t_1 - n,$$

is a solution of (1.8). That is to say

$$U(t) = \sum_{d=1}^{n} R_0(d) \times U(t-d), \forall t \geq t_1.$$

Identification of the components U_i from the values of $t \to N(t)$: Assume that the values of $N(t)$ are given for $t = t_1, \dots, t_1 + n - 1$. We claim that we can compute $U_1, U_2, U_3, \dots, U_n \in \mathbb{C}$. Indeed

$$\begin{array}{llll} N(t_1) & = U_1 \lambda_1^{t_1} & + U_2 \lambda_2^{t_1} & + \dots + U_n \lambda_n^{t_1}, \\ N(t_1+1) & = U_1 \lambda_1^{t_1+1} & + U_2 \lambda_2^{t_1+1} & + \dots + U_n \lambda_n^{t_1+1}, \\ & \vdots & & \\ N(t_1+n-1) & = U_1 \lambda_1^{t_1+n-1} & + U_2 \lambda_2^{t_1+n-1} & + \dots + U_n \lambda_n^{t_1+n-1}, \end{array}$$

can be rewritten as the system

$$\begin{pmatrix} N(t_1) \\ N(t_1+1) \\ \vdots \\ N(t_1+n-1) \end{pmatrix} = \begin{pmatrix} \lambda_1^{t_1} & \lambda_2^{t_1} \cdots\cdots & \lambda_n^{t_1} \\ \lambda_1^{t_1+1} & \lambda_2^{t_1+1} \cdots\cdots & \lambda_n^{t_1+1} \\ \vdots & \vdots & \vdots \\ \lambda_1^{t_1+n-1} & \lambda_2^{t_1+n-1} \cdots\cdots & \lambda_n^{t_1+n-1} \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{pmatrix}. \tag{2.3}$$

The determinant of the above Vandermonde like matrix

$$\det \begin{pmatrix} \lambda_1^{t_1} & \lambda_2^{t_1} \cdots\cdots & \lambda_n^{t_1} \\ \lambda_1^{t_1+1} & \lambda_2^{t_1+1} \cdots\cdots & \lambda_n^{t_1+1} \\ \vdots & \vdots & \vdots \\ \lambda_1^{t_1+n-1} & \lambda_2^{t_1+n-1} \cdots\cdots & \lambda_n^{t_1+n-1} \end{pmatrix} = \lambda_1^{t_1} \lambda_2^{t_2} \dots \lambda_n^{t_n} \prod_{1 \leq i < j \leq n} (\lambda_i - \lambda_j).$$

Therefore under Assumption 2.1 this determinant is non null, and we obtain the following result.

Proposition 2.5. *Let Assumption 2.1 be satisfied. Then we can compute the components $U_1, \dots, U_n$ in function of the given elements of the trajectory $N(t_1), \dots, N(t_1 + n - 1)$ by solving the linear system (2.3), and*

$$\begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{pmatrix} = \begin{pmatrix} \lambda_1^{t_1} & \lambda_2^{t_1} \cdots\cdots & \lambda_n^{t_1} \\ \lambda_1^{t_1+1} & \lambda_2^{t_1+1} \cdots\cdots & \lambda_n^{t_1+1} \\ \vdots & \vdots & \vdots \\ \lambda_1^{t_1+n-1} & \lambda_2^{t_1+n-1} \cdots\cdots & \lambda_n^{t_1+n-1} \end{pmatrix}^{-1} \begin{pmatrix} N(t_1) \\ N(t_1+1) \\ \vdots \\ N(t_1+n-1) \end{pmatrix}.$$

Identification of the component $R_0(d)$ from the λ_i: By assuming that each λ_i is a solution of the characteristic equation (2.2), we obtain

$$\begin{aligned} 1 &= R_0(1)\lambda_1^{-1} + R_0(2)\lambda_1^{-2} + \ldots + R_0(n)\lambda_1^{-n}, \\ 1 &= R_0(1)\lambda_2^{-1} + R_0(2)\lambda_2^{-2} + \ldots + R_0(n)\lambda_2^{-n}, \\ &\vdots \\ 1 &= R_0(1)\lambda_n^{-1} + R_0(2)\lambda_n^{-2} + \ldots + R_0(n)\lambda_n^{-n}, \end{aligned} \tag{2.4}$$

which rewrites in the matrix form as

$$\begin{pmatrix} 1 \\ 1 \\ \vdots \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} \lambda_1^{-1} & \lambda_1^{-2} \cdots\cdots & \lambda_1^{-n} \\ \lambda_2^{-1} & \lambda_2^{-2} \cdots\cdots & \lambda_2^{-n} \\ \vdots & \vdots & \vdots \\ \lambda_n^{-1} & \lambda_n^{-2} \cdots\cdots & \lambda_n^{-n} \end{pmatrix} \begin{pmatrix} R_0(1) \\ R_0(2) \\ \vdots \\ R_0(n) \end{pmatrix}.$$

Under Assumption 1.8 the Vandermonde like matrix

$$\begin{pmatrix} \lambda_1^{-1} & \lambda_1^{-2} \cdots\cdots & \lambda_1^{-n} \\ \lambda_2^{-1} & \lambda_2^{-2} \cdots\cdots & \lambda_2^{-n} \\ \vdots & \vdots & \vdots \\ \lambda_n^{-1} & \lambda_n^{-2} \cdots\cdots & \lambda_n^{-n} \end{pmatrix}$$

is invertible, because

$$\det \begin{pmatrix} \lambda_1^{-1} & \lambda_1^{-2} \cdots\cdots & \lambda_1^{-n} \\ \lambda_2^{-1} & \lambda_2^{-2} \cdots\cdots & \lambda_2^{-n} \\ \vdots & \vdots & \vdots \\ \lambda_n^{-1} & \lambda_n^{-2} \cdots\cdots & \lambda_n^{-n} \end{pmatrix} = \lambda_1^{-1}\lambda_2^{-1}\ldots\lambda_n^{-1} \det \begin{pmatrix} 1 & \lambda_1^{-1} \cdots\cdots & \lambda_1^{-(n-1)} \\ 1 & \lambda_2^{-1} \cdots\cdots & \lambda_2^{-(n-1)} \\ \vdots & \vdots & \vdots \\ 1 & \lambda_n^{-1} \cdots\cdots & \lambda_n^{-(n-1)} \end{pmatrix}$$

hence

$$\det \begin{pmatrix} \lambda_1^{-1} & \lambda_1^{-2} \cdots\cdots & \lambda_1^{-n} \\ \lambda_2^{-1} & \lambda_2^{-2} \cdots\cdots & \lambda_2^{-n} \\ \vdots & \vdots & \vdots \\ \lambda_n^{-1} & \lambda_n^{-2} \cdots\cdots & \lambda_n^{-n} \end{pmatrix} = \lambda_1^{-1}\lambda_2^{-1}\ldots\lambda_n^{-1} \prod_{1 \leq i < j \leq n} \left(\lambda_i^{-1} - \lambda_j^{-1}\right) \neq 0.$$

Therefore, we can compute the component of the map $d \in [1, n] \to R_0(d)$ by solving a linear system involving the eigenvalues of the characteristic equation.

Theorem 2.6. *Let Assumption 2.1 be satisfied. Then the following properties are equivalent*

(i) *The set $\{\lambda_1, \ldots, \lambda_n\}$ is the spectrum of the Leslie matrix L defined in (2.1).*

(ii) *Each element of $\{\lambda_1, \ldots, \lambda_n\}$ satisfies (2.4).*

(iii) *The elements $\{\lambda_1, \ldots, \lambda_n\}$ satisfy*

$$\begin{pmatrix} \lambda_1^{-1} & \lambda_1^{-2} \cdots\cdots & \lambda_1^{-n} \\ \lambda_2^{-1} & \lambda_2^{-2} \cdots\cdots & \lambda_2^{-n} \\ \vdots & \vdots & \vdots \\ \lambda_n^{-1} & \lambda_n^{-2} \cdots\cdots & \lambda_n^{-n} \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} R_0(1) \\ R_0(2) \\ \vdots \\ R_0(n) \end{pmatrix}. \tag{2.5}$$

Definition 2.7. *We will say that L is a* ***Markovian Leslie matrix*** *if all the values $d \in [1,n] \to R_0(d)$ are non negative, and*

$$\sum_{d=1}^{n} R_0(d) = 1.$$

In Figure 4, we plot all the spectrum's location for Markovian Leslie matrices on a mesh. We can observe the changes of location of the spectrum depending of the dimension n. It seems that the spectrum is fielding more and more the unit circle in $\mathbb{C}$ when the dimension increases. We refer to Kirkland [23] for more result going an that direction.

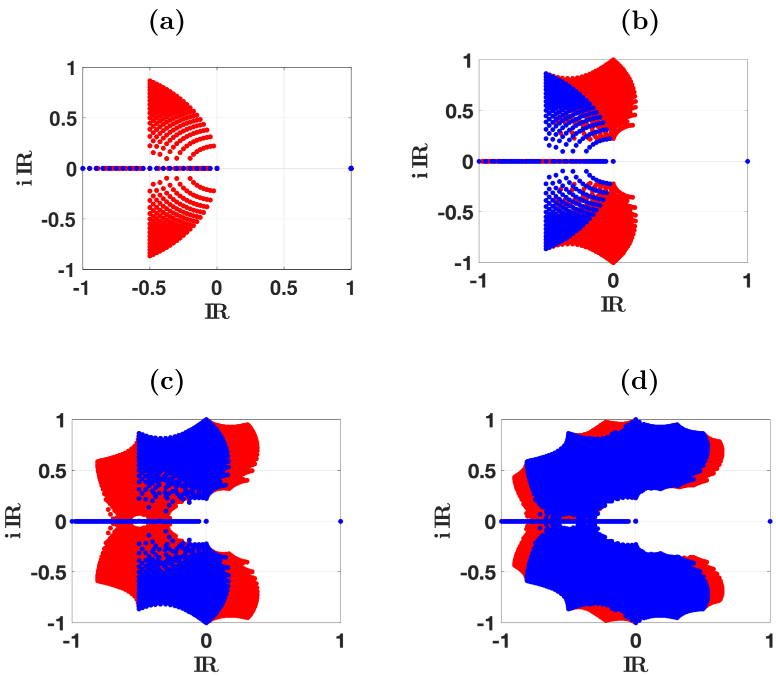

Figure 4: *We plot all the spectrum's location for Markovian Leslie matrices on a mesh whenever $n = 3$ in (a), $n = 4$ in (b), $n = 5$ in (c), and $n = 7$ in (d). Here the dominant eigenvalue is always 1, and we can see the corresponding isolated blue dot. The blue region corresponds to the spectrum of Markovian Leslie matrices whenever $R_0(n) = 0$. The red region corresponds to the spectrum of Markovian Leslie matrices whenever $R_0(n) > 0$.*

Continuous dependency of the component $R_0(d)$ with respect to the λ_i: Define the set $\Omega \subset \mathbb{C}^n$ of all the elements $\Lambda = \{\lambda_1^\star, \ldots, \lambda_n^\star\} \in \mathbb{C}^n$ satisfying Assumption 2.1. For each $\Lambda = \{\lambda_1^\star, \ldots, \lambda_n^\star\} \in \Omega$, we define

$$M(\Lambda) = \begin{pmatrix} \lambda_1^{-1} & \lambda_1^{-2} \cdots\cdots\cdots \lambda_1^{-n} \\ \lambda_2^{-1} & \lambda_2^{-2} \cdots\cdots\cdots \lambda_2^{-n} \\ \vdots & \vdots \qquad\qquad \vdots \\ \lambda_n^{-1} & \lambda_n^{-2} \cdots\cdots\cdots \lambda_n^{-n} \end{pmatrix}, \forall \Lambda = \{\lambda_1, \ldots, \lambda_n\} \in \Omega.$$

Theorem 2.8. *Consider a sequence $\{\Lambda^m = \{\lambda_1^m, \ldots, \lambda_n^m\}\}_{m \geq 0} \subset \Omega$, and a point $\Lambda^\star = \{\lambda_1^\star, \ldots, \lambda_n^\star\} \in \Omega$ (i.e. all satisfying Assumption 2.1). Assume that*

$$\lim_{m \to +\infty} \Lambda^m = \Lambda^\star,$$

then

$$\lim_{m \to +\infty} R_0^m(d) = R_0^\star(d), \forall d = 1, \ldots, n,$$

where

$$R_0^m = M(\Lambda^m)^{-1}\begin{pmatrix}1\\1\\\vdots\\\vdots\\1\end{pmatrix}, \forall m \in \mathbb{N}, \text{ and } R_0^\star = M(\Lambda^\star)^{-1}\begin{pmatrix}1\\1\\\vdots\\\vdots\\1\end{pmatrix}.$$

Proof. We have

$$\begin{pmatrix}1\\1\\\vdots\\\vdots\\1\end{pmatrix} = M(\Lambda^m)R_0^m, \forall n \in \mathbb{N}, \text{ and } \begin{pmatrix}1\\1\\\vdots\\\vdots\\1\end{pmatrix} = M(\Lambda^\star)R_0^\star.$$

Subtracting the two above quantities, we obtain

$$0 = M(\Lambda^m)R_0^m - M(\Lambda^\star)R_0^\star, \tag{2.6}$$

which is also equivalent to

$$0 = M(\Lambda^m)R_0^m - M(\Lambda^\star)\left[R_0^\star - R_0^m\right] - M(\Lambda^\star)R_0^m,$$

hence

$$R_0^\star - R_0^m = M(\Lambda^\star)^{-1}\left[M(\Lambda^m) - M(\Lambda^\star)\right]R_0^m.$$

Setting

$$L_m = M(\Lambda^\star)^{-1}\left[M(\Lambda^m) - M(\Lambda^\star)\right]$$

we obtain

$$R_0^\star - R_0^m = L_m R_0^\star - L_m\left[R_0^\star - R_0^m\right],$$

and since

$$\lim_{m\to+\infty} L_m = 0_{\mathrm{M}_n(\mathbb{C})}$$

we deduce that

$$\|R_0^\star - R_0^m\| \le \|L_m\|_{\mathcal{L}(\mathbb{C}^n)}\|R_0^\star\| + \|L_m\|_{\mathcal{L}(\mathbb{C}^n)}\|R_0^\star - R_0^m\|.$$

Hence for all $m \ge 1$ large enough (i.e. satisfying $\|L_m\|_{\mathcal{L}(\mathbb{C}^n)} < 1$)

$$\|R_0^\star - R_0^m\| \le \frac{\|L_m\|_{\mathcal{L}(\mathbb{C}^n)}}{1 - \|L_m\|_{\mathcal{L}(\mathbb{C}^n)}}\|R_0^\star\|,$$

and the proof is completed. □

2.3 Phenomenological model to fit the cumulative and the daily number of reported case data

Due to Lemma 2.4, we propose the following form the phenomenological to represent the data

$$\mathrm{CR}(t) = \mathrm{CR}_1 e^{\lambda_1 t} + \mathrm{CR}_2 e^{\lambda_2 t} + \mathrm{CR}_3 e^{\lambda_3 t} + \ldots + \mathrm{CR}_n e^{\lambda_n t},$$

where $\mathrm{CR}_1, \ldots, \mathrm{CR}_n \in \mathbb{C}$ are non null, and $\lambda_1 = \alpha_1 + i\omega_1, \ldots, \lambda_n = \alpha_n + i\omega_n \in \mathbb{C}$ are two by two separated.

Remark 2.9. *In the above formula, we allow the constant terms whenever $\lambda_n = 0$.*

Assuming that the unit of time is one day, we have the following relationship between the cumulative number of cases $\mathrm{CR}(t)$ and the daily number of cases $N(t)$

$$\mathrm{CR}(t) = \mathrm{CR}(t_0) + \int_{t_0}^{t} N(\sigma)\, d\sigma.$$

We deduce that the daily number of of reported cases has the following form

$$N(t) = N_1 e^{\lambda_1 t} + N_2 e^{\lambda_2 t} + N_3 e^{\lambda_3 t} + \ldots + N_m e^{\lambda_m t},$$

where $N_1, \ldots, N_n \in \mathbb{C}$ are non null, and $\lambda_1, \ldots, \lambda_m \in \mathbb{C}$ are two by two separated, and $m \leq n$.

Since $N(t)$ is obtained from $\mathrm{CR}(t)$ by computing the first derivative, we have the following relationship

$$N_k = \mathrm{CR}_k \times \lambda_k, \forall k = 1, \ldots, m.$$

Remark 2.10. *For the daily number of cases data $t \to N(t)$ only a few eigenvalues will be tractable. For example Section 3.3 we will consider the following extension*

$$N(t) = N_1 e^{\lambda_1 t} + N_2 e^{\lambda_2 t} + N_3 e^{\lambda_3 t} + N_3 e^{\lambda_3 t} + N_4 e^{\lambda_4 t} + w(t)$$

where $w(t)$ will contain $N_5 e^{\lambda_5 t} + \ldots + N_m e^{\lambda_m t}$ merged together with some random terms.

Remark 2.11 (About the identification of the eigenvalues)**.** *Here we will consider, we assume that the daily number of reported cases has the following form*

$$N(t) = N_1 e^{\lambda_1 t} + N_2 e^{\lambda_2} + N_3 e^{\lambda_3 t} + \ldots + N_m e^{\lambda_m t}, \tag{2.7}$$

where $N_1, \ldots, N_n \in \mathbb{C}$ are non null, and $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ are two by two separated.

If we assume that we know $t \to N(t)$ for all positive integer values $t = 0, 1, 2, \ldots$. Then we can compute the Laplace transform

$$\mathcal{L}(N)(\lambda) = \sum_{t=0}^{\infty} e^{-\lambda t} N(t),$$

which is well defined for all $\lambda \in \mathbb{C}$ such that

$$\mathrm{Re}(\lambda) > \max_{i=1,\ldots,n} \mathrm{Re}(\lambda_i).$$

By using the formula (2.7), we obtain

$$\mathcal{L}(N)(\lambda) = \sum_{p=1}^{m} \frac{N_p}{1 - e^{\lambda_p - \lambda}},$$

whenever $\mathrm{Re}(\lambda) > \max_{i=1,\ldots,n} \mathrm{Re}(\lambda_i)$*.*

Let $k \in \{1, \ldots, m\}$ be an integer such that

$$\mathrm{Re}(\lambda_k) = \max_{i=1,\ldots,n} \mathrm{Re}(\lambda_i),$$

we obtain

$$\lim_{\substack{\lambda \to \lambda_k \\ \mathrm{Re}(\lambda) > \mathrm{Re}(\lambda_k)}} |\mathcal{L}(N)(\lambda)| = +\infty.$$

The Laplace transform could be used to identify the unknown parameters λ_k. Then by combining this idea with linear regression of $t \to e^{\lambda_k t}$, we could identify the parameters N_k. Then step by step compute all the parameters of $N(t)$ in (2.7).

In practice, we only know $t \to N(t)$ on a finite time interval $t = 0, 1, 2, \dots, L$. In that case we can define the Laplace transform has

$$\mathcal{L}(N)(\lambda) = \sum_{t=0}^{L} e^{-\lambda t} N(t)$$

and we have

$$\mathcal{L}(N)(\lambda) = \sum_{p=1}^{m} N_p \frac{1 - e^{(\lambda_p - \lambda)(L+1)}}{1 - e^{(\lambda_p - \lambda)}}.$$

The Laplace transform do not permit to detect the eigenvalues λ_k (we tested without success some examples with values of complex numbers coming from the present article). Identification of the eigenvalues λ_k, whenever $t \to N(t)$ is known only on a finite time interval seams to be an open intriguing question.

We will first approach the data with the following phenomenological model.

Phenomenological model for the cumulative numbers of reported cases with $\lambda > 0$

We start with a first eigenvalue $\lambda = e^{\alpha} > 0$, for some $\alpha \in \mathbb{R}$. The phenomenological model used to fit the cumulative numbers of reported cases has the following form

$$\mathrm{CR}(t) = A e^{\alpha(t - t_0)} + C, \text{ for } t \in [t_0, +\infty), \tag{2.8}$$

where $A \in \mathbb{R}$, $\alpha \in \mathbb{R}$, and $C \in \mathbb{R}$ are real numbers.

For discrete times, that is also equivalent to say that

$$\mathrm{CR}(n) = A\lambda^n + C, \text{ for } n = 0, 1, 2, \dots. \tag{2.9}$$

By computing the first derivative of $t \to \mathrm{CR}(t)$, we obtain a model for the daily number of cases of the following form

$$N(t) = A\,\alpha\, e^{\alpha(t - t_0)}, \text{ for } t \in [t_0, +\infty). \tag{2.10}$$

Once obtained the best fit of the above phenomenological model to the data, we can subtract this model to the data $t \to \mathrm{CR}_{\mathrm{Data}}(t)$, then we obtain a first error

$$\mathrm{Error}(t) = \mathrm{CR}_{\mathrm{Data}}(t) - \mathrm{CR}(t).$$

Next we will approach the error with the following phenomenological model.

Phenomenological model for the cumulative numbers of reported cases with $\lambda \in \mathbb{C}$

Assume that the eigenvalues are two conjugated complex numbers $\lambda = e^{\alpha \pm i\omega} \in \mathbb{C}$, for some $\alpha \in \mathbb{R}$ and $\omega \geq 0$. The phenomenological model used to fit the cumulative numbers of reported cases has the following form

$$\mathrm{CR}(t) = e^{\alpha(t-t_0)} \left[A \cos\left(\omega \left(t - t_0\right)\right) + B \sin\left(\omega \left(t - t_0\right)\right)\right] + C, \text{ for } t \in [t_0, +\infty), \tag{2.11}$$

where $\alpha \in \mathbb{R}$, $A \in \mathbb{R}$, $B \in \mathbb{R}$, $C \in \mathbb{R}$, and $\omega \geq 0$ are four real numbers.

For discrete times, that is also equivalent to say that

$$\mathrm{CR}(n) = \frac{A - iB}{2}\lambda^n + \frac{A + iB}{2}\overline{\lambda}^n + C, \text{ for } n = 0, 1, 2, \ldots. \tag{2.12}$$

By computing the first derivative of $t \to \mathrm{CR}(t)$, we obtain a model for the daily number of cases of the following form

$$N(t) = e^{\alpha(t-t_0)} \left[\widehat{A} \cos\left(\omega \left(t - t_0\right)\right) + \widehat{B} \sin\left(\omega \left(t - t_0\right)\right)\right], \text{ for } t \in [t_0, +\infty), \tag{2.13}$$

where

$$\begin{cases} \widehat{A} = \alpha A + \omega B \\ \widehat{B} = -\omega A + \alpha B \end{cases} \Leftrightarrow \begin{cases} A = \dfrac{\alpha \widehat{A} - \omega \widehat{B}}{\omega^2 + \alpha^2} \\ B = \dfrac{\omega \widehat{A} + \alpha \widehat{B}}{\omega^2 + \alpha^2} \end{cases}. \tag{2.14}$$

Remark 2.12. *When $\omega = 0$ in (2.11), we obtain the previous model (2.8).*

2.4 Cumulative and daily number of reported cases for COVID-19 in Japan

Here we use cumulative numbers of reported cases for COVID-19 in Japan taken from WHO [44]. The data shows a succession of epidemic waves (blue background color regions) followed by endemic periods (yellow background color regions).

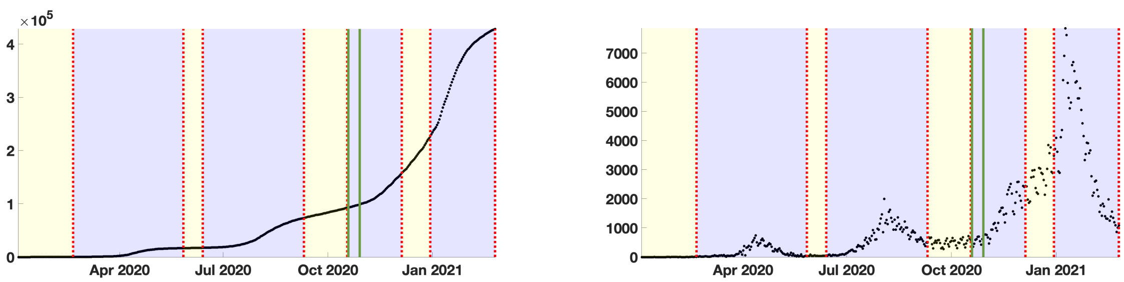

Figure 5: *In this figure, we plot the cumulative number of reported cases (left hand side) and daily number of reported cases (right hand side) for COVID-19 in Japan. Black dots represent the data. The blue background color regions correspond to epidemic phases, and the yellow background color region to endemic phases. The region of interest to apply the method is between October* 19 *and October* 29 2020. *This region is marked with light green vertical lines on the figure.*

3 Results

3.1 Methods applied to one-week data

In this section, we will fit the phenomenological model (2.8) or (2.11) to the cumulative numbers of reported cases presented in the previous subsection. We consider a period of 10 days since the beginning of the third epidemic wave of COVID-19 in Japan. The period goes from October 19 to October 29 2020.

Step 1: In Figure 6, we fit an exponential function (2.8) to the cumulative number of reported cases of COVID-19 in Japan between October 19 and October 29 2020.

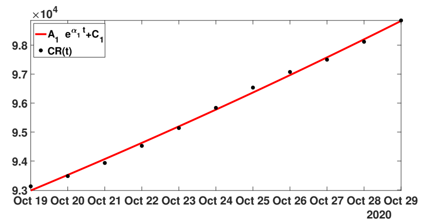

Figure 6: *In this figure, the black dots correspond to the cumulative numbers of reported cases of COVID-19 in Japan between October* 19 *and October* 29 2020 *(black dots). The red curve corresponds to the best fit of model (2.8) to the cumulative number of reported cases.*

In Figure 6, the best fit of model (2.8) is obtained for

$$A_1 = 2.881 \times 10^4, C_1 = 6.4173 \times 10^4, \text{ and } \alpha_1 = 0.0185.$$

Hence

$$\lambda_1 = \exp(\alpha_1) = 1.0187.$$

Step 2: Next, we consider the error left after the previous fit,

$$\text{Error}_1(t) = \text{CR}(t) - \left[A_1 e^{\alpha_1 t} + C_1\right].$$

In Figure 7, we fit the model (2.11) to the first error function $t \to \text{Error}_1(t)$. The period of the error is approximately equal to a week. Actually, the period is equal to $\dfrac{2\pi}{\omega_2} = 6.609$ days. This periodic phenomenon was observed in many countries (see for example [10]).

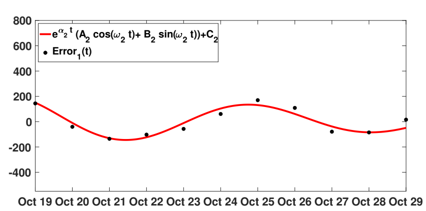

Figure 7: *In this figure, the black dots correspond to the function* $t \to \text{Error}_1(t)$ *from October* 19 *and October* 29 2020 *(black dots). The red curve corresponds to the best fit of model (2.11) to* $\text{Error}_1(t)$.

In Figure 7, the best fit of model (2.11) is obtained for

$$A_2 = 138.1625, B_2 = -127.3613, C_2 = 11.8779, \alpha_2 = -0.0738, \text{ and } \omega_2 = 0.9507.$$

Hence

$$\lambda_2 = \exp(\alpha_2 + i\,\omega_2) = 0.5398 + 0.7560\,i,$$
$$\lambda_3 = \exp(\alpha_2 - i\,\omega_2) = 0.5398 - 0.7560\,i.$$

By using

$$M = \begin{pmatrix} \lambda_1^{-1} & \lambda_1^{-2} & \lambda_1^{-3} \\ \lambda_2^{-1} & \lambda_2^{-2} & \lambda_2^{-3} \\ \lambda_3^{-1} & \lambda_3^{-2} & \lambda_3^{-3} \end{pmatrix},$$

and by (2.5) we obtain

$$\begin{pmatrix} R_0(1) \\ R_0(2) \\ R_0(3) \end{pmatrix} = \begin{pmatrix} 2.0982 \\ -1.9625 \\ 0.8789 \end{pmatrix}. \tag{3.1}$$

Moreover, we obtain

$$\det(M) = 1.7833\,i,$$

therefore the components of M^{-1} are not too large, and the above result should not be too sensitive to the stochastic errors. The main problem in (3.1) is the second component -1.9625 which is not making sense in this context.

3.2 Spectral truncation method applied to one-week data

In the previous subsection, the first two fits make perfect sense. But adding more fits would be questionable because the rest is becoming more and more random after a few steps. We could alternatively continue to fit the rest by using our phenomenological model, which would provide new eigenvalues.

The major problem in the previous section is that when we apply formula (2.5) with all the eigenvalues, we obtain some $R_0(1), \ldots, R_0(n)$ with negative values. Instead here, we increase the dimension n of L, and we use only the eigenvalues $\lambda_1, \lambda_2, \lambda_3$.

3.2.1 Re-normalizing procedure

Assume that $\lambda_1 \neq 1$ then by

$$\overline{N}(t) = \frac{N(t)}{\lambda_1^t} \Leftrightarrow N(t) = \lambda_1^t \times \overline{N}(t)$$

where $t \to N(t)$ is a solution (1.8), we obtain the following normalized equation

$$\lambda_1^t \times \overline{N}(t) = \sum_{d=1}^{n} R_0(d) \times \lambda_1^{t-d} \times \overline{N}(t-d), \forall t \geq t_1,$$

and by dividing the above equation by λ_1^t we obtain

$$\overline{N}(t) = \sum_{d=1}^{n} \overline{R}_0(d)\,\overline{N}(t-d), \forall t \geq t_1.$$

where

$$R_0(d) = \overline{R}_0(d)\lambda_1^d, \forall d = 1, \ldots, n. \tag{3.2}$$

By using the procedure, we can always fix the dominant eigenvalue of L to 1 by imposing that L is Markovian. Then we use the following re-normalizing procedure for the eigenvalues

$$\lambda_1^\star = \lambda_1/\lambda_1 = 1, \lambda_2^\star = \lambda_2/\lambda_1 = 0.5299 + 0.7421\,i, \text{ and } \lambda_3^\star = 0.5299 - 0.7421\,i.$$

In Figure 8, we fit these eigenvalues $\lambda_2^\star$ and $\lambda_3^\star$ with the spectrum of Markovian Leslie matrices L on a mesh. We observe that the fit improves when the dimension of L increases.

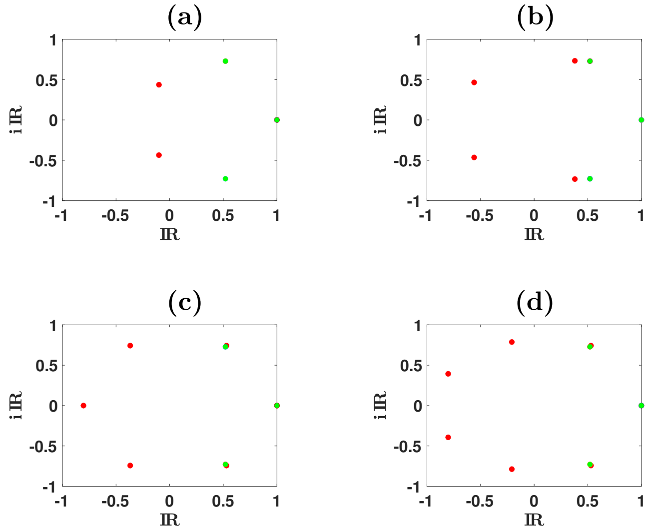

Figure 8: *We plot the spectrum of the Markovian Leslie matrices L (red dots) when* $n = 3, 5, 6, 7$, *(respectively in (a),(b),(c),(d)) giving the best match to the secondary eigenvalues* $\lambda_2^\star$ *and* $\lambda_3^\star$ *(green dots). We observe that the best fit of the two secondary eigenvalues remain faraway from* $\lambda_2^\star$ *and* $\lambda_3^\star$ *for* $n = 3$, *then get closer for* $n = 5$, *and are very close for* $n = 6$ *and* $n = 7$.

In Figure 9, we observe that for $n \in \{3, 5, 6\}$, we deduce that there is a unique set of eigenvalue $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n$ of L (classified with decreasing real part) minimizing the distance $|\lambda_2^\star - \lambda_2|$ and $|\lambda_3^\star - \lambda_3|$. This is no longer true for $n = 7$.

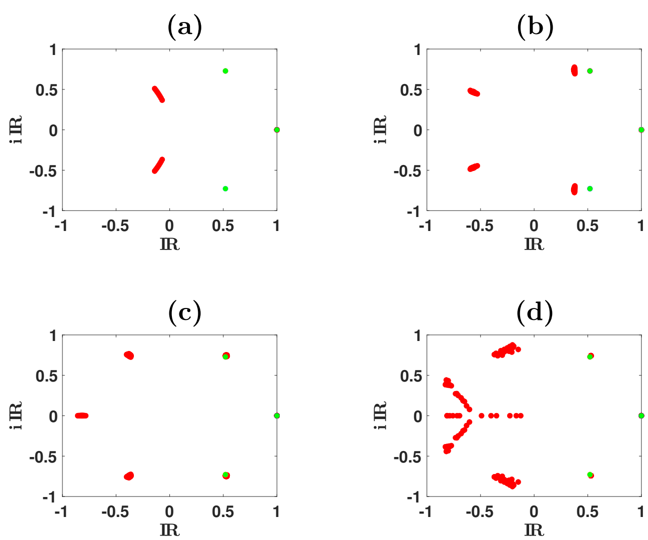

Figure 9: *We plot the spectrum of the Leslie matrix L (red dots) when* $n = 3, 5, 6, 7$, *(respectively in (a),(b),(c),(d)) giving the best match to the secondary eigenvalues* $\lambda_2^\star$ *and* $\lambda_3^\star$ *(green dots). The red dots correspond to the spectrum of L for all the possible matrices L, having their second pair of eigenvalues close to the minimal distance to* $\lambda_2^\star$ *and* $\lambda_3^\star$.

3.2.2 Daily basic reproduction numbers

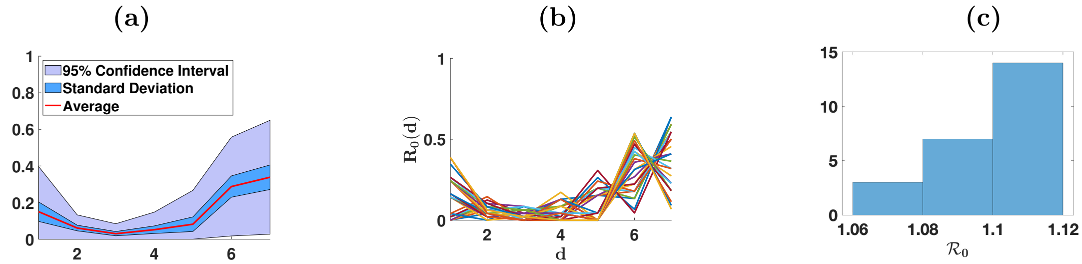

Figure 10: *In this figure we plot $R_0(d)$ which correspond to the multiple close to the best fit to the second eigenvalues whenever $n = 7$. In Figure (a), the average distribution $d \to R_0(d)$ (red curve), standard deviation (blue region), and 95% confidence interval (light blue region). In Figure (b) we plot the 24 distributions $d \to R_0(d)$. In figure (c) make an histogram with the multiple values of $\mathcal{R}_0$. We observe that some of the $d \to R_0(d)$ are similar to the case $n = 6$, with a maximum on day $d = 6$. But on average the maximum value in on day 7.*

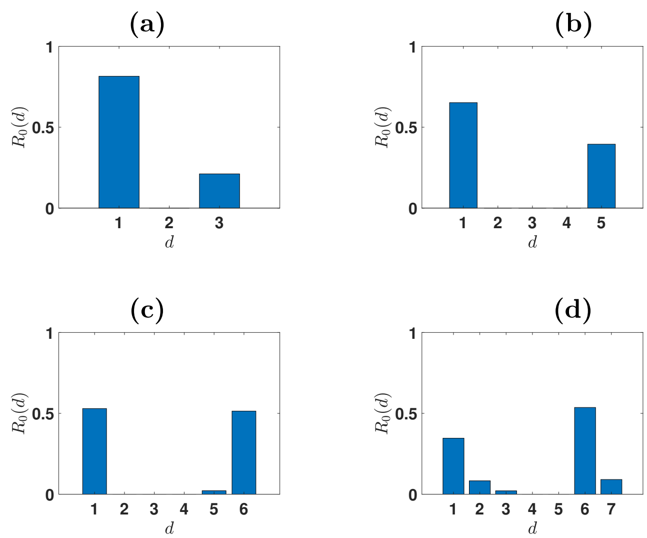

Figure 11: *We plot the daily basic reproduction numbers $R_0(d)$ obtained for $n = 3$ in (a), $n = 5$ in (b), $n = 6$ in (c), and $n = 7$ in (d). The distribution for $n = 7$ corresponds to the red curve in Figure 10.*

n	3	5	6	7
$\mathcal{R}_0$	1.0264	1.0469	1.0658	1.0780

Table 1: *The above reproduction numbers are obtained by using the formula $\sum_{d=1}^{n} R_0(d)$.*

3.2.3 Applying the model to daily number of reported cases

The model used to run the simulations is the following

$$N(t) = \sum_{d=1}^{6} R_0(d) N(t-d), \forall t \geq t_0 + 6, \tag{3.3}$$

and accordingly to the formula (2.10) and (2.13), with the initial condition

$$N(t) = A_1 \ln(\lambda_1) \lambda_1^t + e^{\alpha_2 t}[\widehat{A}_2 \cos(\omega_2 t) + \widehat{B}_2 \sin(\omega_2 t)], \forall t = t_0, t_0+1, \ldots, t_0+5, \tag{3.4}$$

with

$$\widehat{A}_2 = \alpha_2 A_2 + \omega_2 B_2 \text{ and } \widehat{B}_2 = -\omega_2 A_2 + \alpha_2 B_2. \tag{3.5}$$

In (3.3)-(3.5) we use the parameter values estimated in Section 3.1.

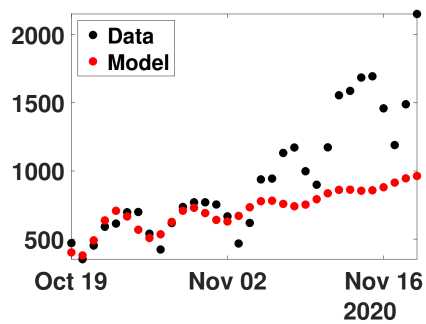

Figure 12: *In this figure we plot the daily number of reported cases data from October* 19 *and November* 19 2020 *(black dots). The model (3.3)-(3.4) with the values of $R_0(d)$ obtained in Figure (11)-(c) (red dots).*

3.3 Extension of the spectral truncation method over one month

In Figures 13 we apply respectively the autocorrelation function (ACF) and partial autocorrelation function (PACF) to the daily number of cases for Japan from October 19 and November 19 2020. It does not look like any standard cases. In the (ACF), we observe the correlation is significant until 7 days, while in the (PACF) it is until 16 days.

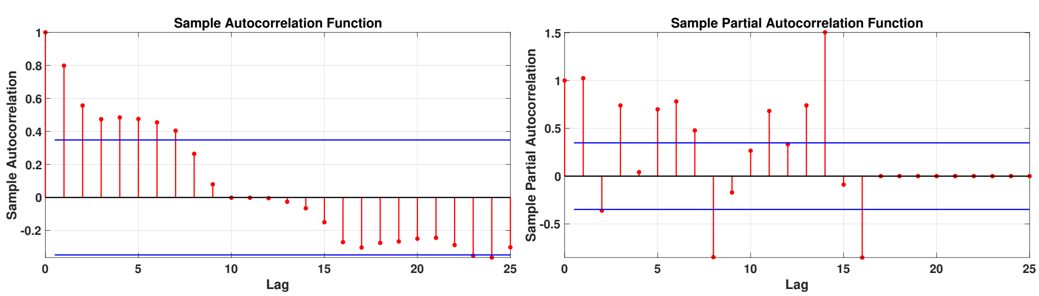

Figure 13: *Autocorrelation Function (ACF) (left hand side) and Partial Autocorrelation Function (PACF) (right hand side) applied to the daily number of cases for Japan from October* 19 *and November* 19 2020*.*

Step 1: In Figure 14, we fit the model

$$\phi_1(t) = A_1 e^{\alpha_1(t-t_0)} + C_1, \tag{3.6}$$

with the cumulative number of reported cases data from October 19 and November 19 2020.

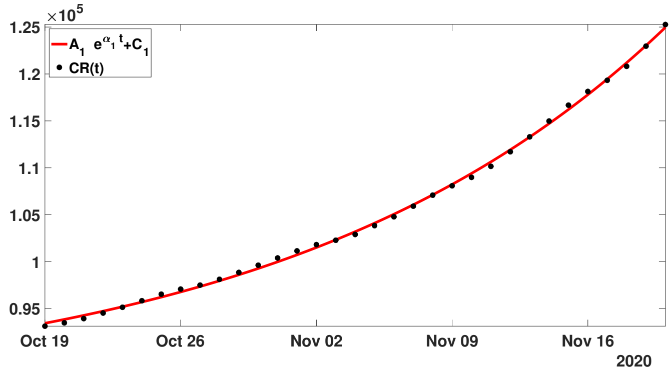

Figure 14: *In this figure, we plot the cumulative number of reported cases data from October* 19 *and November* 19 2020 *(black dots). We plot the best fit of the model (3.6) to the cumulative data (red curve).*

We obtain the following parameter values for the best fit

$$A_1 = 7.9290 \times 10^3, C_1 = 8.5508 \times 10^4, \text{ and } \alpha_1 = 0.0501. \tag{3.7}$$

Step 2: Next we define as before the first error

$$\text{Error}_1(t) = \text{CR}(t) - A_1 e^{\alpha_1 (t-t_0)} + C_1, \tag{3.8}$$

and we fit the $\text{Error}_1(t)$ with the model

$$\begin{aligned} \phi_2(t) \quad &= e^{\alpha_2 (t-t_0)} \left[A_2 \cos\left(\omega_2 \left(t-t_0\right)\right) + B_2 \sin\left(\omega_2 \left(t-t_0\right)\right)\right] \\ &+ e^{\alpha_3 (t-t_0)} \left[A_3 \cos\left(\omega_3 \left(t-t_0\right)\right) + B_3 \sin\left(\omega_3 \left(t-t_0\right)\right)\right] + C_2. \end{aligned} \tag{3.9}$$

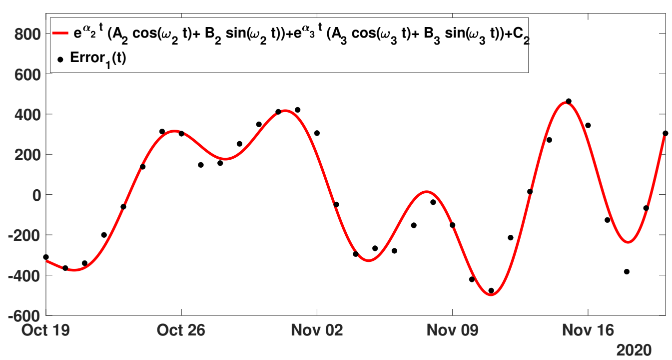

Figure 15: *In this figure, we plot the cumulative number of reported cases data from October* 19 *and November* 19 2020 *(black dots). We plot the best fit of the model (3.9) to the cumulative data (red curve).*

The parameters of the phenomenological model $\phi_2(t)$ obtained for the best fit are the following

$$A_2 = 55.2075, B_2 = -84.4842, A_3 = -391.5688, B_3 = 88.7878, C_2 = 7.6835, \tag{3.10}$$

and

$$\alpha_2 = 0.0501, \omega_2 = 0.9080, \alpha_3 = -0.0198, \omega_3 = 0.2958. \tag{3.11}$$

The periods associated to ω_2 and ω_3 are

$$P_2 = \frac{2\pi}{\omega_2} = 6.9198 \text{ days, and } P_3 = \frac{2\pi}{\omega_3} = 21.2425 \text{ days.}$$

These periods are close multiple of 7 days.

Remark 3.1. *It is important to note that the period of P_3 of 21 days is difficult to explain mechanically, but this value is the smallest value giving the best fit to the data. We tried to impose some upper bounds smaller than 21 days. In such a case P_3 is always replaced by the upper bound. This is true for all constraints less that 21 days, and for each constraint larger than 22 days, we obtain $P_3 = 21.24$ days.*

Remark 3.2. *It is important to note that $\alpha_1 = \alpha_2$. That is because during the fit we impose that $\alpha_2 \leq \alpha_1$ and $\alpha_3 \leq \alpha_1$. That is the condition coming from the Perron Frobenius theorem, in order to obtain*

$$|\lambda_2| \leq |\lambda_1| \text{ and } |\lambda_3| \leq |\lambda_1|.$$

This condition is coming from the fact that λ_1 must be the spectral radius of L and λ_2, λ_3 belong to the circle centered at 0 and with radius spectrum of L (i.e. with a modulus less or equal to λ_1).

Eigenvalues associated to the model $\phi_1(t)$ and $\phi_2(t)$: The first eigenvalue is

$$\lambda_1 = e^{\alpha_1} = 1.0514.$$

The second pair of complex conjugated eigenvalues is

$$\lambda_2 = e^{\alpha_2}\left[\cos(\omega_2) + i\sin(\omega_2)\right] = 0.6470 + 0.8288\, i,$$

and the modulus of λ_2 is

$$|\lambda_2| = e^{\alpha_2} = e^{\alpha_1} = \lambda_1 = 1.0514.$$

The forth eigenvalue is

$$\lambda_4 = e^{\alpha_3}\left[\cos(\omega_3) + i\sin(\omega_3)\right] = 0.9386 + 0.2865\, i.$$

and its modulus us

$$|\lambda_4| = e^{\alpha_3} = 0.9804 < 1.0514.$$

Using λ_2 and λ_4 as an estimator: Next we consider all the matrices L in which the component $R_0(d)$ are replaced by $\overline{R}_0(d)$, and we assume that

$$\sum_{d=1}^{n} \overline{R}_0(d) = 1.$$

The dominant eigenvalue of L is 1, and we look for matrices such that the second eigenvalue of L is close to

$$\lambda_2^\star = \lambda_2/\lambda_1,$$

and the fourth eigenvalue of L is close to

$$\lambda_4^\star = \lambda_4/\lambda_1.$$

For realizing this approach, we minimize the

$$\chi(L) = \max\left(d\left(\lambda_2^\star, \sigma(L)\right), d\left(\lambda_4^\star, \sigma(L)\right)\right)$$

where

$$d\left(\lambda_2^\star, \sigma(L)\right) = \min_{\lambda \in \sigma(L)} |\lambda_2^\star - \lambda|, \text{ and } d\left(\lambda_4^\star, \sigma(L)\right) = \min_{\lambda \in \sigma(L)} |\lambda_4^\star - \lambda|,$$

where $\sigma(L)$ is the set of all eigenvalues of L.

In the Figure 16 we consider the $d \to \overline{R}_0(d)$ such that the corresponding max $\widehat{L}$ satisfies

$$\chi\left(L(\overline{R}_0)\right) \leq \inf_{\overline{R}_0 \geq 0 : \sum \overline{R}_0(d)=1} \chi\left(L(\overline{R}_0)\right) + 10^{-2}.$$

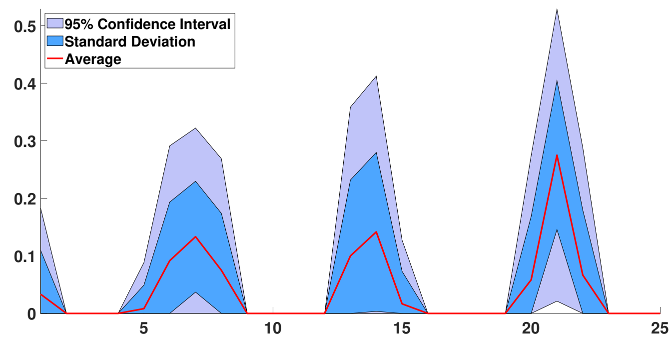

Figure 16: *In this figure, we consider the case $n = 25$. We plot the distributions of daily basic reproduction numbers $d \to \overline{R}_0(d)$ corresponding to the distributions having some secondary eigenvalues and fourth eigenvalues at a distance less than 10^{-2} to the best match. The red curve is the average distribution $d \to \overline{R}_0(d)$. The blue region corresponds to the standard deviation around the mean distribution. The light blue region corresponds to the 95% confidence interval.*

We define

$$R_0(d) = \overline{R}_0(d)\lambda_1^d, \forall d = 1, \ldots, n. \tag{3.12}$$

In Figure 17, we obtain a good description of the dynamic of infection at the individual level that confirms the one obtained over shorter periods. As expected, the average patient first looses its ability to transmit the pathogen, and after decreasing by day 1 to day 4, $R_0(d)$ increases between day 4 and day 7. Day 7 is a maximum. After the day 7, $R_0(d)$ decay until day 9. Then a second pic arises, with a maximum on the day 14. We could explain this second peak by supposing that an important transmission of pathogen still exists from day 12 to day 16. We also obtain a third from day 19 to 23 with a maximum value on day 21.

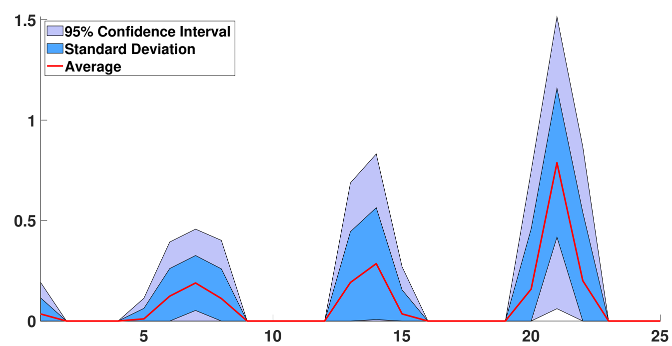

Figure 17: *In this figure, we consider two cases $n = 25$. We plot the distributions of daily basic reproduction numbers $d \to R_0(d) = \overline{R}_0(d)\lambda_1^d$, where $\overline{R}_0(d)$ is the red curve in Figure 16.*

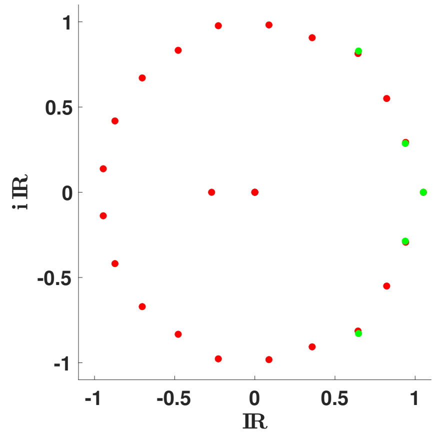

Figure 18: *In this figure, we consider the case* $n = 25$. *We plot the spectrum of the Leslie matrix* L *(red dots) when* $n = 25$ *and* $d \to \overline{R}_0(d)$ *corresponds to the average distribution (i.e. the red curve in Figure 16).*

The basic reproduction number is obtained by summing the

$$\mathcal{R}_0 = \sum_{d=1}^{n} R_0(d).$$

We obtain the mean value of the daily reproduction numbers (red curve in the Figure 17)

$$\mathcal{R}_0 = 2.1316.$$

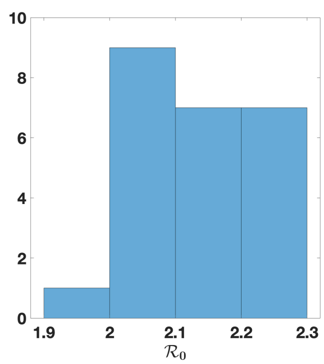

Figure 19: *In this figure, we consider the case* $n = 25$. *In this figure, we plot a histogram for the values of the basic reproduction number obtained by summing the distributions* $d \to R_0(d)$ *from Figure 17.*

Next we consider

$$N(t) = \sum_{d=1}^{25} R_0(d) N(t-d), \forall t \geq t_0 + 25, \tag{3.13}$$

and accordingly to the formula (2.10) and (2.13), with the initial condition for $t = t_0, t_0 + 1, \ldots, t_0 + 25$, we have

$$N(t) = A_1 \ln(\lambda_1) \lambda_1^t + e^{\alpha_2 t}[\widehat{A}_2 \cos(\omega_2 t) + \widehat{B}_2 \sin(\omega_2 t)] + e^{\alpha_3 t}[\widehat{A}_3 \cos(\omega_3 t) + \widehat{B}_3 \sin(\omega_3 t)], \tag{3.14}$$

with

$$\widehat{A}_2 = \alpha_2 A_2 + \omega_2 B_2, \widehat{B}_2 = -\omega_2 A_2 + \alpha_2 B_2, \widehat{A}_3 = \alpha_3 A_3 + \omega_3 B_3 \text{ and } \widehat{B}_3 = -\omega_3 A_3 + \alpha_3 B_3. \quad (3.15)$$

In (3.3)-(3.5) we use the parameter values estimated in Section 3.1.

In Figure 20, we see the mean distribution $d \to R_0(d)$ permits to produce oscillations around the tendency for the daily number of cases. It is important to note that without the third pic in Figure 17 we do not obtain such a good correspondence between the model and the data.

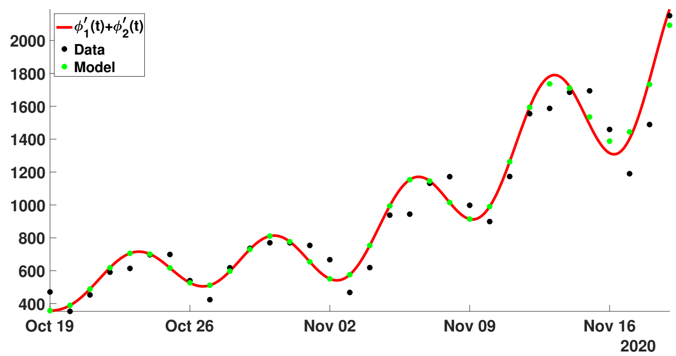

Figure 20: *In this figure we plot the daily number of reported cases data from October* 19 *and November* 19 2020 *(black dots). The red curve corresponds to* $\phi_1' + \phi_2'$*, and the green dots correspond (3.13)-(3.14) whenever* $R_0(d)$ *coming from the is the average distribution (i.e. the red curve in Figure 16). We observe a very good match between the green dots and the red curve (the phenomenological model).*

4 Discussion

In this article, we start by investigating the connection between a signal decomposed into a sum of damped or amplified oscillations and a renewal equation. Namely, we connect the daily number of reported written as

$$N(t) = N_1 e^{\alpha_1 t} \left[\cos(\omega_1 t) + i \sin(\omega_1 t)\right] + \ldots + N_n e^{\alpha_n t} \left[\cos(\omega_n t) + i \sin(\omega_n t)\right], \ \forall t \geq t_1 - n,$$

with the renewal equation

$$N(t) = \sum_{d=1}^{n} R_0(d) \times N(t-d), \forall t \geq t_1.$$

In the context of epidemic time series, a spectral method usually refers to the Fourier decomposition of a periodic signal. In the present paper, the data are not periodic and are composed of an exponential function (Malthusian growth) perturbed with some damped oscillating functions. So we use complex numbers with non-null real parts. We refer to Cazelles et al. [4] for more results about time series.

4.1 Data over one week

We can notice on Figure 10, Figure 11, and Table 1 that the daily reproduction number as well as the instantaneous reproduction number are estimated. The instantaneous (or effective) reproduction number $R_e(t)$ [7], [36] estimated by [45], which equals 1.1 at the 19th of October 2020, the best fit corresponding to $n = 7$ days (see (c) in Figure 10). This value of the duration of the contagiousness period is close to the values 6 or 7 days are close to the values estimated from the virulence measured in [22] [31] [20]. In Figure 11, we always obtain a U-shape distribution for the curve of daily reproduction numbers. This corresponds to the biphasic form of the virulence already observed in respiratory viroses, such as influenza as recalled in the Introduction.

This temporal behavior of the contagiousness can correspond to the evolution of contagious symptoms like cough or spitting, which diminish during the innate immune response, followed by a comeback of the symptoms before the adaptive immune response (whenever the innate defense has been overcome by the virus). If the innate cellular immunity has been not sufficient for eliminating the virus, the viral load anew increases causing a reappearance of the symptoms before the adaptive immunity (cellular and humoral) occurs, which results in a transient decrease in contagiousness between the two immunologic phases. The medical recommendations are, in case of U-shape of the contagiousness, never to take a transient improvement for a permanent disappearance of the symptoms and to stay at home to avoid a bacterial secondary infection possibly fatal.

The estimation of the daily reproduction numbers in COVID-19 outbreak constitutes an important issue. At the public health level, to publish only the sum of the daily reproduction numbers, that is to say the basic reproduction number R0 or the effective reproduction number Re, could suffice for controlling and managing the behavior of a whole population with mitigation or vaccination measures. At the individual level, it is important to know the existence of a minimum of the daily reproduction numbers, which generally corresponds to a temporary clinical improvement, after a partial success of the innate immune defense: this makes it possible to advice the patient to continue to respect his own isolation, prevention and therapy choices (depending on his vaccination state) even if this transient clinical improvement has occurred. The present methodology allows also to estimate both the individual contagiousness duration in a dedicated age class and also its seasonal variations, which is crucial for optimizing the benefit-risk decisions of the public and individual health policies.

4.2 Data over one month

Over one month, we obtain a daily reproduction number with three pics. Each pic is centered respectively on 7 days, 14 days, and 21 days. These quantities coincide with the period 7 days and 21 days obtained in Figure 15 in fitting the first error when we subtract the exponential growing first fit to the cumulative data. As far as we understand the problem, that is the period of 21 days in the data, which induces the third pic. This third pic is very suspicious. Nevertheless, the data lead us to such a shape for the daily reproductive number. We also tried to run Figure 20 without the third pic, and we obtained a really bad fit to the data, while with this third pic, the fit is really good. One may also note that 21 days is insignificant for the ACF and the PACF in Figures 13.

Several possibilities exist to explain this strange shape for the daily reproduction number using the data over one month. One possible explanation is that the Japanese population should be subdivided into several groups having very different infection dynamics (at the level of a single patient). Here we have in mind the patient with a short infection period but high transmissibility (super spreaders) versus the patient with a long infection period with mild symptoms.

We suspect that such a shape for the daily reproduction number could be attributed to the time since infection to report a case. The daily number of reported cases would obtained from $N(t)$ the daily number of new infected by using the following model

$$D(t) = f \sum_{d=1}^{q} K(d) N(t-d),$$

where $f \in [0,1]$ is the fraction reported, and $K(d) \geq 0$ is the probability to report a case after d days. Therefore we must have

$$\sum_{d=1}^{q} K(d) = 1.$$

4.3 Perspectives and conclusion

In the present paper, we only consider the Japanese data in the exponential phase of the third epidemic wave. Such a method should be applied to several countries for each epidemic wave to obtain a more systematic study. For the moment, over one month, we got a shape for the daily reproduction number that follows the data very well. But we are suspicious about the third pic. We suspect that the default of our analysis is coming from the model itself. Such a question has been recently studied by Ioannidis and his collaborators in [18], and we believe that we are facing such modeling difficulties.

Appendix

A About $\text{Error}_2(t)$ in Section 3.3

In Figure 21 we observe that average of $\text{Error}_2(t) = \text{Error}_1(t) - \phi_2(t)$ is close to 0, but its histogram do not have the shape of a normal distribution. So, there might be some residual information in $\text{Error}_2(t)$.

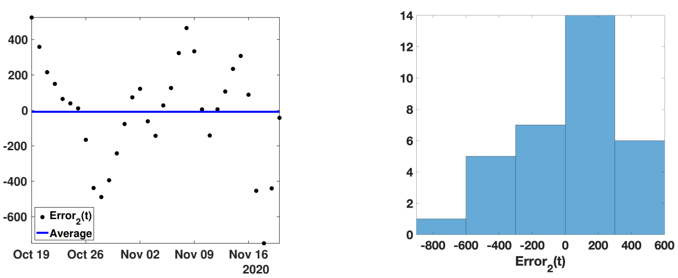

Figure 21: *In this figure we plot* $\text{Error}_2(t)$.

References

[1] L. Alvarez, M. Colom, J. D. Morel, and J. M. Morel, Computing the daily reproduction number of COVID-19 by inverting the renewal equation using a variational technique. *Proceedings of the National Academy of Sciences*, **118(50)** (2021).

[2] L. Alvarez, J-D. Morel, and J-M. Morel, Modeling COVID-19 incidence by the renewal equation after removal of administrative bias and noise, *Biology* **11(4), 540** (2022).

[3] A. Bakhta, T. Boiveau, Y. Maday, and O. Mula, Epidemiological forecasting with model reduction of compartmental models. application to the COVID-19 pandemic. *Biology*, **10(1)**, 22 (2020).

[4] B. Cazelles, M. Chavez, G. C. D. Magny, J. F. Guégan, and S. Hales, Time-dependent spectral analysis of epidemiological time-series with wavelets. *Journal of the Royal Society Interface*, **4(15)** (2007), 625-636.

[5] K.S. Chan, and H. Tong, A note on certain integral equations associated with non-linear time series analysis. *Probab. Th. Rel. Fields* **73** (1986), 153-158.

[6] D.L. Chao, M. E. Halloran, V. J. Obenchain, and Jr, I. M. Longini, FluTE, a publicly available stochastic influenza epidemic simulation model. PLoS computational biology, **6(1)**, e1000656 (2010).

[7] A. Cori, N.M. Ferguson, C. Fraser, and S. Cauchemez. A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am. J. Epidemiol.*, **178(9)** (2013), 1505-1512.

[8] J. Demongeot, Q. Griette and P. Magal, SI epidemic model applied to COVID-19 data in mainland China. *Royal Society Open Science* 7:201878 (2020).

[9] J. Demongeot, Q. Griette, Y. Maday, and P. Magal, Kermack-McKendrick model with age of infection starting from a single or multiple cohorts of infected patients, arXiv preprint arXiv:2205.15634.

[10] J. Demongeot, K. Oshinubi, M. Rachdi, H. Seligmann, F. Thuderoz and J. Waku, Estimation of Daily Reproduction Rates in COVID-19 Outbreak. *Computation*, **9**, 109 (2021).

[11] A. Ducrot, Q. Griette, Z. Liu, & P. Magal, *Differential Equations and Population Dynamics I: Introductory Approaches*. Springer Nature (2022).

[12] L. Euler, Recherches générales sur la mortalité et la multiplication du genre humain. *Mémoires de l'académie des sciences de Berlin*, **16** (1767), 144-164.

[13] Fisher, R.A. The Wave of Advance of Advantageous Genes. Annals of *Eugenics*, **7** (1937), 353–369.

[14] Q. Griette, J. Demongeot and P. Magal, What can we learn from COVID-19 data by using epidemic models with unidentified infectious cases? *Mathematical Biosciences and Engineering*, **19(1)** (2021), 537-594.

[15] Q. Griette, J. Demongeot and P. Magal, A robust phenomenological approach to investigate COVID-19 data for France, *Mathematics in Applied Sciences and Engineering*, **2(3)** (2021), 149-218.

[16] Q. Griette and P. Magal, Clarifying predictions for COVID-19 from testing data: the example of New York State, *Infectious Disease Modelling*, **6** (2021), 273-283.

[17] Hahn, G.M. Mammalian cell populations. Math. Biosci., 6: 295–315 (1970).

[18] Ioannidis, J. P., Cripps, S., and Tanner, M. A., Forecasting for COVID-19 has failed. *International Journal of Forecasting*, **38(2)** (2022), 423-438.

[19] Y. Itoh et al. Emergence of H7N9 Influenza A Virus Resistant to Neuraminidase Inhibitors in Nonhuman Primates. *Antimicrobial Agents and Chemotherapy* **59** (2015), 4962-4973.

[20] H. Kawasuji et al. Transmissibility of COVID-19 depends on the viral load around onset in adult and symptomatic patients. *PLoS ONE*, **15: e0243597** (2020).

[21] W. O. Kermack and A. G. McKendrick, Contributions to the mathematical theory of epidemics: II, *Proc. R. Soc. Lond. Ser. B*, **138** (1932), 55-83.

[22] Kim, S.E.; Jeong, H.S.; Yu, Y.; Shin, S. U.; Kim, S.; Oh, T.H.; Kim, U.J.; Kang, S.J.; Jang, H.C.; Jung, S. I.; Park, K.H. Viral kinetics of SARS-CoV-2 in asymptomatic carriers and presymptomatic patients. *Int. J. Infectious Diseases* **95** (2020), 441-443.

[23] S. Kirkland, On the spectrum of a Leslie matrix with a near-periodic fecundity pattern. *Linear algebra and its applications*, **178** (1993), 261-279.

[24] Lambert, J.H. Beytrage Zum Gebrauche Der Mathematik Und Deren Anwendung. Verlage des Buchladens der Realschule, Berlin (1765-72).

[25] Leslie, P.H. On the use of matrices in certain population mathematics. Biometrika, 33:183–212 (1945).

[26] K.S. Lim, and H. Tong, A statistical approach to difference-delay equation modelling in ecology – two case studies. *Journal of Time Series Analysis*, **4** (1983), 239-267.

[27] A. J. Lotka, Relation between birth rates and death rates. *Science*, 26: (1907), 121–130.

[28] T.R. Malthus, An Essay on the Principle of Population as It Affects the Future Improvement of Society, with Remarks on the Speculations of Mr. Godwin, M. Condorcet, and Other Writers. J. Johnson, London (1798).

[29] H. Nishiura, Time variations in the transmissibility of pandemic influenza in Prussia, Germany, from 1918-19. Theor. Biol. Med. Model. 4, 20 (2007).

[30] H. Nishiura, G. Chowell, "The effective reproduction number as a prelude to statistical estimation of time-dependent epidemic trends" in Mathematical and Statistical Estimation Approaches in Epidemiology, G. Chowell, J. M. Hyman, L. M. A. Bettencourt, C. Castillo-Chavez, Eds. (Springer, Dordrecht, Netherlands, 2009), pp. 103-121.

[31] Y. Pan, D. Zhang, P. Yang, L.L.M. Poon, and Q. Wang, Viral load of SARS-CoV-2 in clinical samples. *Lancet Infect. Dis.*, **20** (2020), 411-412.

[32] Priestley, M. B. Evolutionary spectra and non-stationary processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 27(2) (1965), 204-229.

[33] Priestley, M.B. *Spectral Analysis and Time Series*. Academic Press, London (1981).

[34] Ramsay, J.O. Monotone Regression Splines in Action. *Statistical Science*, 3 (1988), 425-441 .

[35] Ramsay, J.; Hooker, G. Dynamic Data Analysis: Modeling Data with Differential Equations. Springer, New York (2017).

[36] Scire et al. Reproductive number of the COVID-19 epidemic in Switzerland with a focus on the Cantons of Basel-Stadt and Basel-Landschaft. *Swiss Med. Wkly.*, 150(19-20):w20271, 2020.

[37] Tong, H. Some comments on the Canadian lynx data. Journal of the Royal Statistical Society: Series A, 140: 432-436 (1977).

[38] Tong, H. Non-linear time series: a dynamical system approach. Oxford University Press, Oxford (1990).

[39] Tuan, P.D. The estimation of parameters for autoregressive moving average models. Journal of Time Series Analysis, 5: 53-68 (1984).

[40] Verhulst, P.F. Notice sur la loi que la population poursuit dans son accroissement. Correspondance mathématique et physique, 10: 113-121 (1838).

[41] J. Waku, K. Oshinubi, and J. Demongeot, Maximal reproduction number estimation and identification of transmission rate from the first inflection point of new infectious cases waves: COVID-19 outbreak example. *Mathematics and Computers in Simulation*, **198** (2022), 47-64.

[42] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. MIT Press, Cambridge (Mass.) (1949).

[43] Q. Yao, and H. Tong, Quantifying the Influence of Initial Values on Non-Linear Prediction. *Journal of the Royal Statistical Society: Series B*, **56** (1994), 701-725.

[44] Data from WHO. Accessed: July 20, 2022. https://COVID19.who.int/WHO-COVID-19-global-data.csv

[45] Powered by the Institute of Global Health, Faculty of Medicine, University of Geneva and the Swiss Data Science Center, ETH Zürich-EPFL. Accessed: July 20, 2022. https://renkulab.shinyapps.io/COVID-19-Epidemic-Forecasting/_w_850fb011/?tab=jhu_pred&country=Japan

[46] P. Whittle, Hypothesis Testing in Time Series Analysis. *Almquist and Wicksell*, (1951).

[47] P. Whittle, *Prediction and Regulation.* English Universities Press, (1963).