Epidemic with age of infection

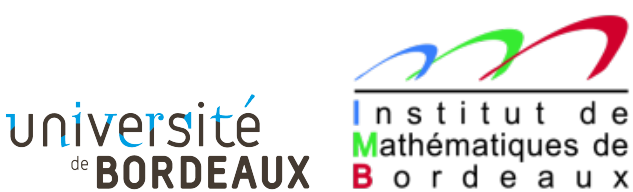
Pierre Magal

University of Bordeaux, France

pierre.magal@u-bordeaux.fr

Summer Bootcamp of Infectious Disease Modelling 2023

Kyoto University, August, 2023.



Abstract

The infectiousness of infected individuals is known to depend on the time since the individual was infected. That is called the age of infection. Here we study the parameter identifiability of the Kermack- McKendrick model with the age of infection. By considering a single cohort of individuals, we show that the daily reproduction number is the solution of a Volterra integral equation that depends on the flow of newly infected individuals. We test the consistency of the method by running some deterministic and stochastic numerical simulations. Then we use the outcome of the simulations as data to reconstruct the basic reproduction number. Finally, we apply our method to a dataset for SARS-CoV-1 with detailed information on a single cluster of patients. We stress the necessity of taking into account the initial data in the analysis to ensure the identifiability of the problem.

Another essential aspect of understanding an epidemic is the contact between individuals which mainly depends on their location at home or work. To investigate such a question, we will discuss a new model including the people at home, commuting, and working in a city.

Continuous time model

Recall that the **age of infection** a is the time since individuals become infected. The major difficulty in matching the data and the Kermack-McKendrick model with age of infection is to identify:

- 1) the initial distribution of infected individuals with respect to the age of infection;
- 2) the **daily reproduction number** $R_0(a)$ which is the reproduction number at the age of infection a.

We can decompose the daily reproduction number as follows

$$R_0(a) = \underbrace{\tau_0}_{\text{(A)}} \times \underbrace{S_0}_{\text{(B)}} \times \underbrace{\beta(a)}_{\text{(C)}} \times \underbrace{e^{-\nu a}}_{\text{(D)}},$$

where

- (A) τ_0 is the transmission rate at time t_0 (we assume the transmission rate to be constant during the period where $R_0(a)$ is evaluated).
- (B) S_0 is the average number of susceptible individuals at time t_0 with which an infected person may come into contact (we assume the number of susceptible individuals to be constant during the period where $R_0(a)$ is evaluated).
- (C) $\beta(a)$ is the probability to be infectious (i.e. capable to transmit the pathogen) for an infected individual with age of infection a days.
- (D) $e^{-\nu a}$ is the probability for an infected individual with age of infection a days to remain infected.

Then the basic reproduction number (i.e. the number of secondary cases produced by a single infected individual) is given by

$$R_0 = \int_0^\infty R_0(a)da.$$

Here we partly solve the problem of finding the initial distribution of infected by assuming that we start the epidemic at time t_0 with a **single cohort** of I_0 new infected patients.

That is, the epidemic starts with I_0 infected patients all with age of infection $a = 0$.

The case of an epidemic starting from a single infected patient (usually called the patient 0) corresponds to the case $I_0 = 1$.

This is a common assumption in epidemiology. But the model is not known.

U or M shape distribution

Here aims to investigate the shape of the distribution $d \mapsto R_0(d)$ from the data.

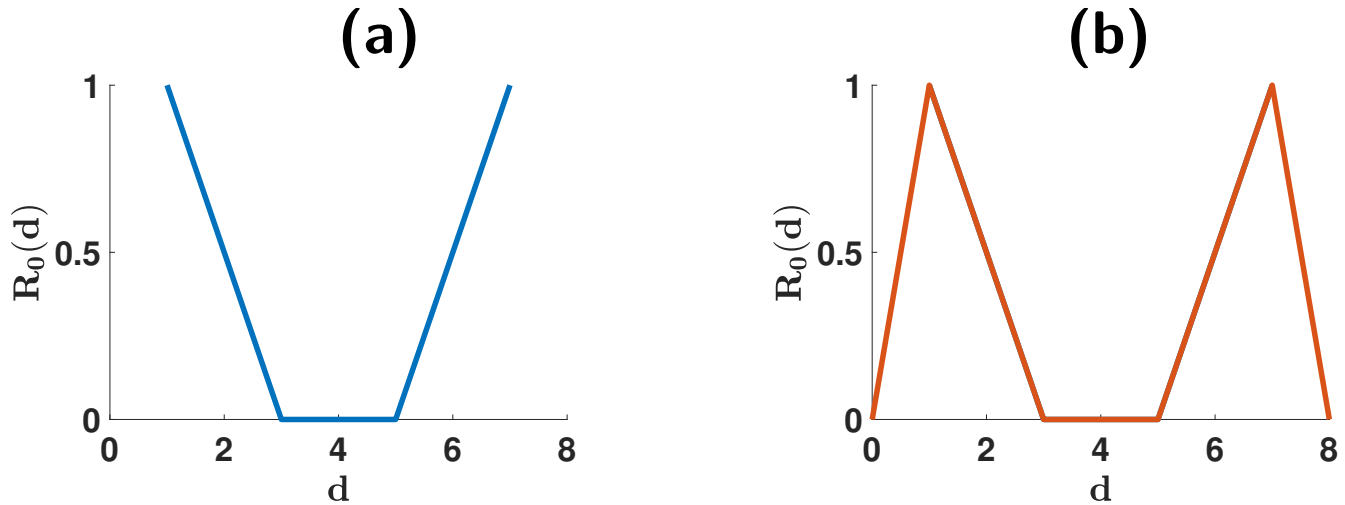


Figure: *In this figure, we illustrate the notion of* U *shape distribution in (a) and* M *shape distribution in (b). Recall that* $R_0(d)$ *represents the ability of patients to transmit the pathogen after* d *days since they got infected. The* U *shape or* M *shape distribution means that patients 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.*

U or M shape distribution

The U or M shape distribution are well known in the context of influenza [1] [2].

[1] 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).

[2] Y. Itoh, S. Shichinohe, M. Nakayama, M. Igarashi, A. Ishii, H. Ishigaki, H. Ishida, N. Kitagawa, T. Sasamura, M. Shiohara, M. Doi, H. Tsuchiya, S. Nakamura, M. Okamatsu, Y. Sakoda, H. Kida, K. Ogasawara, Emergence of H7N9 Influenza A Virus Resistant to Neuraminidase Inhibitors in Nonhuman Primates. *Antimicrobial Agents and Chemotherapy* **59** (2015), 4962-4973.

Viral load in COVID-19 real patients [3]. Here we present some figures reflecting patients' viral load for COVID-19.

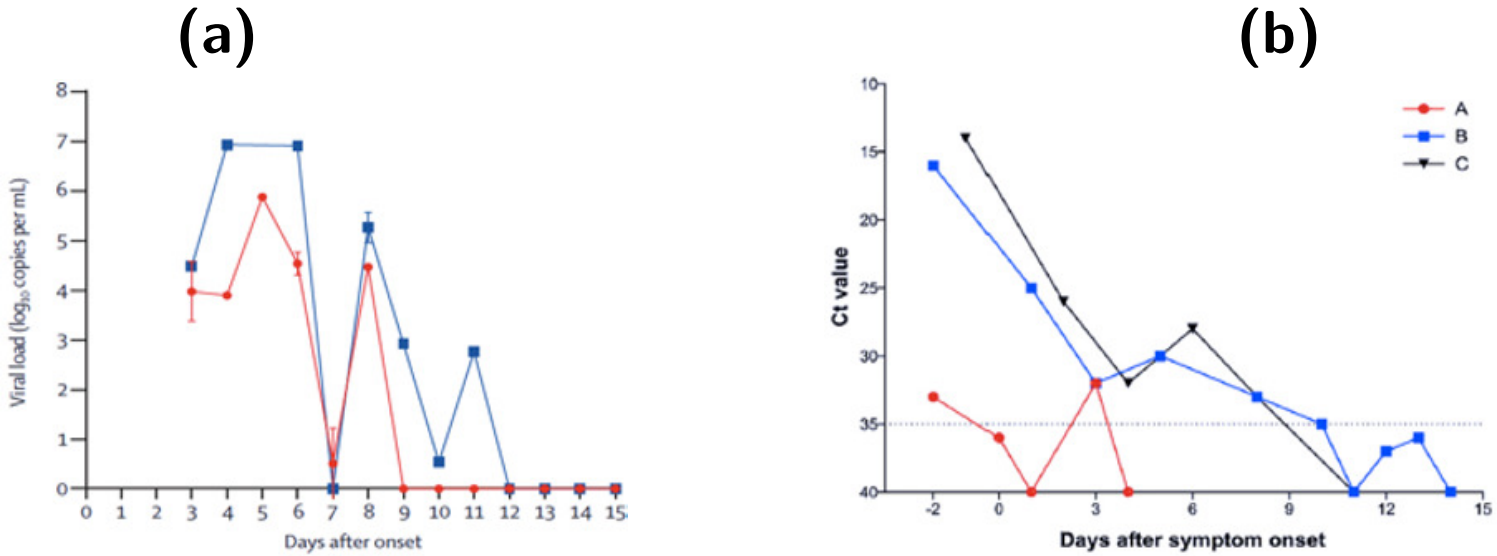


Figure: *In figure (a) the red curve corresponds to the throat swab and the blue curve corresponds to the sputum. In figure (b) the curves correspond to several patients (A), (B), and (C).*

[3] **Y. Pan, D. Zhang, P. Yang, L. L. M. Poon, and Q. Wang** (2020), Viral load of SARS-CoV-2 in clinical samples. *The Lancet Infectious Diseases*, **20(4)**, 411-412.

Kermack-McKendrick model with age of infection

Partial differential equation formulation of the model

The age of infection a is the time since individuals become infected. Let $a \to i(t,a)$ be the distribution of population of *infected individuals* at time t (with respect to a the age of infection). The term distribution of population means that the integral

$$\int_{a_1}^{a_2} i(t,a)da$$

is the number of infected at time t with infection age between a_1 and a_2. Therefore the total number of infected individuals at time t is

$$I(t) = \int_0^{+\infty} i(t,a)da.$$

Let $\beta(a) \in [0,1]$ be the probability to be contagious or infectious (i.e. capable to transmit the pathogen) at the age of infection a. The quantity $\beta(a)$ can be interpreted as the fraction of infected individuals with age of infection a that are infectious. Then the total number of *contagious individuals* (or also called *infectious individuals*) (i.e., the individuals capable of transmitting the pathogen) at time t is

$$C(t) = \int_0^{+\infty} \beta(a) i(t,a) da.$$

The model of Kermack-McKendrick [4] with age of infection is the following, for each $t \geq t_0$

$$\begin{cases} S'(t) = -\tau(t)\, S(t) \displaystyle\int_0^{+\infty} \beta(a)\, i(t,a) da, \\ \partial_t i + \partial_a i = -\nu\, i(t,a), \text{ for } a \geq 0, \\ i(t,0) = \tau(t)\, S(t) \displaystyle\int_0^{+\infty} \beta(a)\, i(t,a) da, \end{cases} \tag{1}$$

this system is supplemented by initial data

$$S(t_0) = S_0 \geq 0, \text{ and } i(t_0,a) = i_0(a) \in L^1_+(0,\infty). \tag{2}$$

where $L^1_+(0,\infty)$ is the positive cone of non-negative integral function.

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

In the model, $S(t)$ is the number of susceptible individuals at time t, and $t \to \tau(t)$ is the transmission rate at time t, and $\nu \geq 0$ is the rate at which individuals die or recover.
The time changes of the transmission rate $\tau(t)$ is the combination of three factors.

- The coefficient of virulence, linked to the infectious agent. The coefficient of virulence may change over time due to mutations of the pathogen.
- The coefficient of susceptibility, linked to the host. This two first factors are all summarized into the probability of transmission.
- The number of contacts per unit of time between individuals (this number is directly connected to the mitigation measures). Here, the parameter ν is assumed to be independent of the age of infection a. This is a simplifying assumption to improve the readability of the paper. The parameter ν combines both the specific fatality rate and the recovery rate.

The number of infected satisfies the folloiwng equation

$$I'(t) = \underbrace{\tau(t)\, S(t) \int_0^{+\infty} \beta(a)\, i(t,a) da}_{\text{(I)}} - \underbrace{\int_0^{+\infty} \nu\, i(t,a) da}_{\text{(II)}},$$

where (I) is the flow of new infected, and (II) is the flow of individuals who die or recover.

We make the following assumption.

Assumption 0.1

We assume that

(i) *The transmission rate* $t \to \tau(t)$ *is a continuous function (**no jumps**)*

$$\tau(t) \geq 0, \forall t \geq t_0.$$

(ii) $\beta(a)$ *the probability to be infectious at the age of infection* a *is a continuous function (**no jumps**)*

$$0 \leq \beta(a) \leq 1, \forall a \geq 0.$$

Kermack-McKendrick model with age of infection

Volterra integral equation formulation of the model

In the model (1), the quantity

$$N(t) := \tau(t)\, S(t) \int_0^{+\infty} \beta(a)\, i(t,a) da, \tag{3}$$

is the flow of new infected individuals at time t. By using the S-equation in system (1), we obtain

$$S(t) = S_0 - \int_{t_0}^{t} N(\sigma) d\sigma, \forall t \geq t_0. \tag{4}$$

By integrating the second equation of system (1) along the characteristics, we obtain

$$i(t,a) = \begin{cases} e^{-\nu\,(t-t_0)}\, i_0\left(a-(t-t_0)\right), & \text{if } a \geq t-t_0, \\ e^{-\nu\, a} N(t-a), & \text{if } t-t_0 \geq a. \end{cases} \tag{5}$$

By using (5), we deduce that $t \to N(t)$ satisfies the following Volterra integral equation

$$\begin{aligned} N(t) = & \underbrace{\tau(t)\, S(t) \int_{t-t_0}^{+\infty} \beta(a)\, e^{-\nu\,(t-t_0)} i_0\,(a-(t-t_0))\, da}_{\text{(I)}} \\ & \underbrace{+\tau(t)\, S(t) \int_{0}^{t-t_0} \beta(a) e^{-\nu\, a} N(t-a)\, da,}_{\text{(II)}} \end{aligned} \tag{6}$$

where

- (I) is the flow of new infected individuals at time t produced by the **infected individuals already present on day** t_0;
- (II) is the flow of new infected individuals at time t produced by the **new infected individuals since day** t_0.

By using equations (6), we can summarize the epidemic model (1), by saying that $t \to N(t)$ is the unique continuous map satisfying

$$N(t) = \tau(t)\, S(t) \left[\Lambda(t) + \int_0^{t-t_0} \beta(a) e^{-\nu\, a} N(t-a)\, da \right], \forall t \geq t_0, \tag{7}$$

where

$$S(t) = S_0 - \int_{t_0}^{t} N(\sigma) d\sigma, \forall t \geq t_0, \tag{8}$$

and

$$\Lambda(t) := e^{-\nu\,(t-t_0)} \int_{t-t_0}^{+\infty} \beta(a)\, i_0\,(a-(t-t_0))\, da, \forall t \geq t_0. \tag{9}$$

The function $\Lambda(t)$ is the number of infectious individuals (capable to transmit the pathogen) at time t among the infected individuals already present at time t_0.

The function $t \to \Lambda(t)$ plays a fundamental role in solving the Volterra equation. Indeed, the quantity

$$\int_{t_1}^{t_2} \tau(\sigma)\, S(\sigma)\Lambda(\sigma)d\sigma,$$

is the number of infected produced between the instants t_1 and t_2 by the infected already present at time t_0.

So, for example, if no new infected are produced by the infected already present at time t_0, that is if $\Lambda(t) = 0, \forall t \geq t_0$, then there will be no new infected at all after the time t_0, that is

$$N(t) = 0, \forall t \geq t_0.$$

The function $t \to \Lambda(t)$ can be regarded as the **initial condition (or initial distribution)** for the Volterra integral equation.

Kermack-McKendrick model with age of infection

Connection with the standard SI model

In the case of the standard SI model, which is

$$\begin{cases} S'(t) = -\tau(t)S(t)I(t), \\ I'(t) = \tau(t)S(t)I(t) - \nu I(t), \end{cases} \quad \text{for } t \geq t_0.$$

by applying the variation of constant formula to I-equation, we obtain

$$I(t) = e^{-\nu(t-t_0)}I_0 + \int_{t_0}^{t} e^{-\nu(t-s)}N(s)ds.$$

Replacing $I(t)$ by the above formula in $N(t) = \tau(t)S(t)I(t)$, we obtain

$$N(t) = \underbrace{\tau(t)\, S(t)\, e^{-\nu\,(t-t_0)} I_0}_{\text{(I)}} + \underbrace{\tau(t)\, S(t) \int_0^{t-t_0} e^{-\nu\, a} N(t-a)\, da}_{\text{(II)}}. \quad (10)$$

In the special case, we have $\beta(a) = 1$, for all $a \geq 0$, and the remaining number of infectious individuals at time t from the infectious individuals that was present at time t_0 is

$$\Lambda(t) = e^{-\nu\,(t-t_0)} I_0.$$

Kermack-McKendrick model with age of infection

Connecting the data and the model

The data are represented by the function $t \to \mathrm{CR}(t)$ which is the cumulative number of reported cases at time t. We propose as a model that the flow of reported cases is a fraction $0 \leq f \leq 1$ of the flow of recovering individuals, that is

$$\mathrm{CR}'(t) = f\,\nu \int_0^{+\infty} i(t,a)da. \tag{11}$$

By using (5), we can compute the number of infected at time t. That is

$$\int_0^{+\infty} i(t,a)da = e^{-\nu\,(t-t_0)} I_0 + \int_0^{t-t_0} e^{-\nu\,a} N(t-a)\,da, \tag{12}$$

where

$$I_0 = \int_0^{+\infty} i_0(a)da$$

is the total number of infected at time t_0.

By using equations (11) and (12), we obtain

$$\mathrm{CR}'(t) = f\nu \left[e^{-\nu\,(t-t_0)} I_0 + \int_0^{t-t_0} e^{-\nu\,a} N(t-a) da \right],$$

or equivalently (by using the change of variable $\sigma = t - a$)

$$\mathrm{CR}'(t) = f\nu \left[e^{-\nu\,(t-t_0)} I_0 + \int_{t_0}^{t} e^{-\nu\,(t-\sigma)} N(\sigma) d\sigma \right].$$

By choosing $t = t_0$ we obtain

$$I_0 = \frac{\mathrm{CR}'(t_0)}{f\nu},$$

and

$$\int_{t_0}^{t} e^{\nu\sigma} N(\sigma) d\sigma = \frac{e^{\nu\,t}\mathrm{CR}'(t)}{f\nu} - e^{\nu\,t_0} I_0,$$

and by differentiating both sides of the above equation, we obtain

$$e^{\nu t} N(t) = \frac{\nu e^{\nu\,t}\mathrm{CR}'(t) + e^{\nu\,t}\mathrm{CR}''(t)}{f\nu}.$$

Therefore we obtain the following connection between the data and the model.

Connection between the data and the model

Let $t \to \mathrm{CR}(t)$ be the cumulative number of reported cases. Then the initial number of infected is given by

$$I_0 = \frac{\mathrm{CR}'(t_0)}{f\nu}, \tag{13}$$

and the flow of new infected individuals $N(t)$ at time t is given by

$$N(t) = \frac{\nu \mathrm{CR}'(t) + \mathrm{CR}''(t)}{f\nu}, \forall t \geq t_0. \tag{14}$$

In practice it is possible but not easy to have a reliable evaluation of $t \to \mathrm{CR}'(t)$ and especially $t \to \mathrm{CR}''(t)$. This problem was considered by using some averaging (or phenomenological models) procedure of the reported sanitary data [5,6,7,8].

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

[6] **J. Demongeot, Q. Griette, and P. Magal** (2020), SI epidemic model applied to COVID-19 data in mainland China. *R. Soc. Open Sci.* **7.12**, 201878.

[7] **Q. Griette, J. Demongeot, and P. Magal** (2021), A robust phenomenological approach to investigate COVID-19 data for France . *Math. Appl. Sci. Eng.*, 2021.

[8] **Q. Griette, J. Demongeot, and P. Magal** (2022), What can we learn from COVID-19 data by using epidemic models with unidentified infectious cases? *Mathematical Biosciences and Engineering*, **19.1**, 537-594.

Kermack-McKendrick model starting from a single and multiple cohorts of infected patients

The major difficulty to compare the model with the data is to identify the functions $a \to i_0(a)$ and $a \to \beta(a)$. To simplify the discussion, let us consider the model at the early stage of the epidemic. When the epidemic just starts we can assume that the transmission rate $t \to \tau(t)$ remains constant, and the number of susceptible individuals $t \to S(t)$ is constant and equal to S_0. Under such a simplifying assumption the Volterra equation becomes

$$N(t) = \tau\, S_0 \left[\Lambda(t) + \int_0^{t-t_0} \beta(a) e^{-\nu\, a} N(t-a)\, da\right], \forall t \geq t_0. \qquad (15)$$

Kermack-McKendrick model starting from a single and multiple cohorts of infected patients

A single cohort initial distribution for the PDE model

In order to understand the mathematical concept of Dirac mass centered at 0, we first consider an approximation by an exponential law

$$i_0(a) = I_0 \, \kappa \, e^{-\kappa a}, \tag{16}$$

with mean and standard deviation equal to $1/\kappa$. Then a Dirac mass centered at age 0 can be understood as the limit of such a distribution when κ goes to $+\infty$. The limit needs some explanations. Recall that

$$\int_{a_1}^{a_2} i_0(a) da = I_0 \, \left[e^{-\kappa a_1} - e^{-\kappa a_2} \right],$$

is the initial number of infected individuals with infection age a in between a_1 and a_2 at time $t = 0$.

We deduce that

$$\lim_{\kappa \to \infty} \int_{a_1}^{a_2} i_0(a) da = \begin{cases} 0, & \text{if } a_2 > a_1 > 0, \\ I_0, & \text{if } a_2 > a_1 = 0. \end{cases}$$

That is to say that, when κ tends to $+\infty$, the initial distribution of population $i_0(a)$ is approaching the case where all the infected individuals at time t_0 have the same age of infection $a = 0$.
For short, we write

$$i_0(a) = I_0 \, \delta_0(a),$$

where $\delta_0(a)$ is called the Dirac mass centered at age 0.

Kermack-McKendrick model starting from a single and multiple cohorts of infected patients

A single cohort initial distribution for the Volterra integral equation

Recall that

$$\Lambda(t) = e^{-\nu\,(t-t_0)} \int_0^{+\infty} \beta\,(a + (t-t_0))\; i_0\,(a)\,da,$$

so when $i_0(a)$ is replaced by $I_0\,\kappa\,e^{-\kappa a}$ we obtain

$$\Lambda_\kappa(t) := I_0\,e^{-\nu\,(t-t_0)} \int_0^{+\infty} \beta\,(a + (t-t_0))\; \kappa\,e^{-\kappa a} da.$$

From now on, every function that depends on κ will be indexed by κ.
In order to derive the Kermack-McKendrick model with Dirac mass initial distribution as limit, we first need the following result. The proof of the following can be found in the supplementary material.

Lemma 0.2

Let Assumption 0.1 be satisfied. Then we have

$$\lim_{\kappa\to\infty} \Lambda_\kappa(t) = I_0 e^{-\nu\,(t-t_0)}\beta\,(t-t_0)\,,$$

where the limit is uniform in $t \geq t_0$.

The initial condition $\Lambda(t)$ of the Volterra integral equation becomes at the limit

$$\Lambda(t) = I_0 e^{-\nu\,(t-t_0)} \beta\,(t-t_0)\,.$$

One may observe that the above limit can be obtained for many types of approximation of the Dirac mass at centered 0 (probability distribution on $(0,+\infty)$). So formula (16) can be replaced by another formula.

Kermack-McKendrick model starting from a single and multiple cohorts of infected patients

A single cohort Volterra integral equation model

Define

$$\Gamma(a) = e^{-\nu\, a}\beta\left(a\right), \forall a \geq 0. \tag{17}$$

Then by using (6), the Kermack-McKendrick model can be reformulated for $t \geq t_0$, as the following system

$$N_\kappa(t) = \tau(t) S_\kappa(t) \left[\Lambda_\kappa(t) + \int_0^{t-t_0} \Gamma(a)\, N_\kappa(t-a) da\right],$$

where $\Lambda_\kappa(t)$ is defined above, and

$$S_\kappa(t) = S_0 - \int_{t_0}^{t} N_\kappa(\sigma) d\sigma.$$

By taking the limit when $\kappa \to +\infty$, we obtain the model starting from a single cohort of infected.

Kermack-McKendrick model starting from a single cohort of infected

Assume that the initial distribution of infected only contains a single cohort composed of I_0 individuals all with age of infection $a = 0$ at time t_0. Then the flow of new infected $t \to N(t)$ is the unique continuous solution of the Volterra integral equation

$$N(t) = \tau(t)S(t)\left[I_0 \times \Gamma\left(t - t_0\right) + \int_0^{t-t_0} \Gamma(a)\, N(t-a)da\right], \forall t \geq t_0, \tag{18}$$

where

$$\Gamma(a) = e^{-\nu\, a}\beta\left(a\right), \forall a \geq 0.$$

and

$$S(t) = S_0 - \int_{t_0}^{t} N(\sigma)d\sigma, \forall t \geq t_0.$$

- The PDE model starting from Dirac mass has no continuous time solution.
- So the PDE model does not extend to initial distributions in the space of measures.
- But the Volterra equation integral equation can be extended to the case of initial distributions, which are measures.

In the case of multiple cohorts, the initial distribution becomes

$$i_0(a) = I_0^1\,\delta_{a_1}(a) + \ldots + I_0^n\,\delta_{a_n}(a),$$

where $a_1 < a_2 < \ldots < a_n$ are the ages of infection for each cohort at time t_0, and I_0^j is the number of infected in the j^{th}-cohort at time t_0. By analogy to the case of a single cohort, we can approach the initial condition as follows

$$\Lambda_\kappa(t) := \sum_{j=1}^{n} I_0^j\, e^{-\nu\,(t-t_0)} \int_0^{+\infty} \beta\,(a + (t-t_0))\;\kappa\, e^{-\kappa(a-a_j)}\,\mathbb{1}\,(a-a_j)\,da,$$

Kermack-McKendrick model starting from multiple cohorts of infected

Assume that the initial distribution of infected consists in $n \geq 1$ cohorts of infected with age of infection $a_1 < a_2 < \ldots < a_n$ at time t_0. That is

$$i_0(a) = I_0^1 \, \delta_{a_1}(a) + \ldots + I_0^n \, \delta_{a_n}(a).$$

where I_0^j is the number of infected in the j^{th}-cohort at time t_0. Then the flow of infected $t \to N(t)$ satisfies the following Volterra integral equation

$$N(t) = \tau(t) S(t) \left[\sum_{j=1}^{n} \Gamma(t - t_0 + a_j) \frac{I_0^j}{e^{-\nu a_j}} + \int_0^{t-t_0} \Gamma(a) \, N(t-a) da \right],$$

where

$$S(t) = S_0 - \int_{t_0}^{t} N(\sigma) d\sigma, \forall t \geq t_0.$$

Kermack-McKendrick model starting from a single and multiple cohorts of infected patients

Basic reproduction number

Define the **daily reproduction numbers**

$$R_0(a) = \tau \times S_0 \times \Gamma(a) = \tau \times S_0 \times \beta\,(a) \times e^{-\nu\, a}, \forall a \geq 0.$$

Assume that the transmission $t \to S(t)$ and $t \to \tau(t)$ are constant functions, we obtain

$$N(t) = I_0 \times R_0(t - t_0) + \int_0^{t-t_0} R_0(a)\, N(t-a)da, \;\; \forall t \geq t_0.$$

By using the change of variable $s = t - t_0$,

$$N(s + t_0) = I_0 \times R_0(s) + \int_0^{s} R_0(a)\, N(s + t_0 - a)da, \;\; \forall s \geq 0.$$

Replacing the notation s by t, and define

$$N_{t_0}(t) = N(t + t_0), \forall t \geq 0,$$

the equation (18) becomes

$$N_{t_0}(t) = \left[I_0 \times R_0(t) + \int_0^{t} R_0(a)\, N_{t_0}(t-a)da \right], \forall t \geq 0. \qquad (19)$$

Convolution of functions defined only for positive numbers

We define the **convolution** between two functions $a \in [0, \infty) \mapsto U(a)$ and $a \in [0, \infty) \mapsto V(a)$,

$$(U * V)(t) = \int_0^t U(a)V(t-a)da = \int_0^t U(t-a)V(a)da.$$

$$\left(U^{*(2)}\right)(t) = (U * U)(t),$$

and for each integer $n \geq 3$,

$$\begin{aligned} \left(U^{*(n)}\right)(t) &= \left(U^{*(n-1)} * U\right)(t) \\ &= \left(U * U^{*(n-1)}\right)(t) \\ &= \underbrace{(U * U * \ldots * U)}_{\text{n times}}(t). \end{aligned}$$

By replacing $N_{t_0}(t)$ by the right hand side of the previous Volterra equation in the integral term of (19) we obtain by induction

$$
\begin{aligned}
N_{t_0}(t) = \quad & I_0 R_0(t) \\
& + I_0 (R_0 * R_0)(t) \\
& + I_0 (R_0 * R_0 * R_0)(t) \\
& \vdots \\
& + I_0 \underbrace{(R_0 * R_0 * \ldots * R_0)}_{\text{n times}}(t) \\
& \vdots
\end{aligned}
$$

We can interpret the N-equation concretely as follows

$$N_{t_0}(t) = \underbrace{I_0\, R_0(t)}_{\text{Flow of new infected produced by the first generation of infected individuals}}$$

$$+ \underbrace{I_0 \left(R_0^{*(2)}\right)(t)}_{\text{Flow of infected produced the second generation of infected individuals}}$$

$$+ \underbrace{I_0 \left(R_0^{*(3)}\right)(t)}_{\text{Flow of infected produced by the third generation of infected individuals}}$$

$$+$$

$$\vdots$$

$$+ \underbrace{I_0 \left(R_0^{*(n)}\right)(t)}_{\text{Flow of infected produced by the } n^{th} \text{ generation of infected individuals}}$$

$$+$$

$$\vdots$$

Basic reproduction number

The total number of the first generation of new infected produced by a single infected patient with age of infection $a = 0$ at time $t = t_0$ is called the **basic reproduction number**. That is

$$R_0 = \int_0^\infty R_0(a)\, da.$$

The flow of the first generation of new infected produced by a single infected patient who has been infected for a days is called the **daily reproduction numbers**. When the time unit is one day, the function $R_0(a)$ is also the average daily number of case produced by a single patient at the age of infection a.

Computing the age dependent reproduction number $\Gamma(a)$ from the data

Computing $\Gamma(a)$ from the data

Assume in addition that the parameters $t_0, S_0 > 0, I_0 > 0, \nu > 0$, and the function $t \to \tau(t)$ are known. Then the function $t \to \Gamma(t)$ can be obtained from the flow of new infected $t \to N(t)$, as the unique solution of the Volterra integral equation

$$\Gamma(t - t_0) = \frac{1}{I_0} \left(\frac{N(t)}{\tau(t) S(t)} - \int_0^{t-t_0} \Gamma(a)\, N(t-a) da \right), \forall t \geq t_0,$$

where

$$S(t) = S_0 - \int_{t_0}^{t} N(\sigma) d\sigma, \forall t \geq t_0.$$

Day by day Kermack-McKendrick model with age of infection

Day by day single cohort model and daily basic reproduction number

Assume that $t \to \tau(t)$ equal τ_0, and $t \to S(t)$ is constant equal to S_0. Assume that the epidemic starts at time t_0 with a cohort of I_0 new infected patients (i.e. with age of infection $a = 0$). The model with a single cohort of infected becomes a discrete Volterra equation

$$N(t) = \left[R_0(t - t_0) \times I_0 + \sum_{d=1}^{t-t_0} R_0(d-1) \times N(t-d) \right], \forall t \geq t_0.$$

Define $a = t - t_0$, the age since t_0, we obtain an equation for the daily reproduction number

$$R_0(a) = \frac{N(t_0 + a)}{I_0} - \frac{1}{I_0} \sum_{d=1}^{a} R_0(d-1) \times N(t_0 + a - d), \forall a \geq 0.$$

This gives

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

We obtain

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

which is equivalent to

$$
\begin{aligned}
R_0(0) &= N(t_0)/I_0, \\
R_0(1) &= \Big(N(t_0+1) - R_0(0)N(t_0)\Big)/I_0, \\
R_0(2) &= \Big(N(t_0+2) - R_0(1)N(t_0) - R_0(0)N(t_0+1)\Big)/I_0, \\
R_0(3) &= \Big(N(t_0+3) - R_0(2)N(t_0) - R_0(1)N(t_0+1) - R_0(2)N(t_0+2)\Big)/I_0, \\
&\vdots
\end{aligned}
$$

Numerical simulations

In the simulations, the unit of time is one day, and we fix

$$S_0 = 10^7 = 10\,000\,000,\ 1/\nu = 9 \text{ days, and } R_0 = 1.1.$$

For each function $\beta(a)$ described below, the parameter τ is obtained numerically by using the following formula

$$\tau = \frac{R_0}{S_0 \int_0^\infty \beta(a) e^{-\nu a} da},$$

where the integral is computed by using the Simpson integration method.

Numerical simulations

Stochastic simulations: Individual Based Model (IBM)

In order to estimate the uncertainty expected in real datasets, we use stochastic simulations that reproduce the first stages of the epidemic in finite populations.
We consider a population composed of a finite number

$$N = S_0 + I_0$$

of individuals. We start the simulation a time $t = 0$ with

- $S_0 \in \mathbb{N}$ susceptible individuals
  and
- $I_0 \in \mathbb{N}$ infected individuals all with age of infection $a = 0$.

For each infected individuals we also compute the time spent in the I-compartment which follows an exponential law with parameters $1/\nu$.

The principles of the simulations are as follows:

1. Individuals meet at random at rate $\tau > 0$. In other words, each pair of individual in the population has a contact which occurs at a time following an exponential law of average $1/\tau$.
2. When a contact occurs between an infected individual of age a and a susceptible individual, the contact results in a newly infected individual of age 0 with probability $\beta(a)$.
3. When the infection occurs, the newly infected individual is assigned a duration of infection which follows an exponential law of rate ν. Therefore individuals stay infected on average for a duration of $1/\nu$.
4. The age of all individuals is updated at fixed intervals of time of size Δt. Simultaneously the life-span of each infected invidual is decreased by Δt and individuals whose life-span has become negative are removed from the system.

The MATLAB code of the IBM is available online at: https://github.com/romainvieme/2022-kermack-mckendrick-single-cohort.

Numerical simulations

Numerical evidence of the convergence of the IBM to the deterministic model

It is common to see biphasic flu clinically: after incubation of one day, there is a high fever, then a drop in temperature before rising again, hence the term "V" fever[9]. Such a biphasic contagiousness is also observed in COVID-19. The viral load in throat swab and sputum has been measured for COVID-19 patients, which leads to biphasic contagiousness[10,11] .

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

[10] **Y. Pan, D. Zhang, P. Yang, L. L. M. Poon, and Q. Wang** (2020), Viral load of SARS-CoV-2 in clinical samples. *The Lancet Infectious Diseases*, **20(4)**, 411-412.

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

To cover these type of infectious diseases, we introduce the following form for the probability to be infectious

$$\beta(a) = 0.5 \times 4q \left\{ (a - a_0) \left(1 - q(a - a_0)\right) \right\}^+ + 4q \left\{ (a - pa_0) \left(1 - q(a - pa_0)\right) \right\}^+, \tag{20}$$

with $a_0 = 3$ days, $p = 2.5,$ and $q = 0.3$.

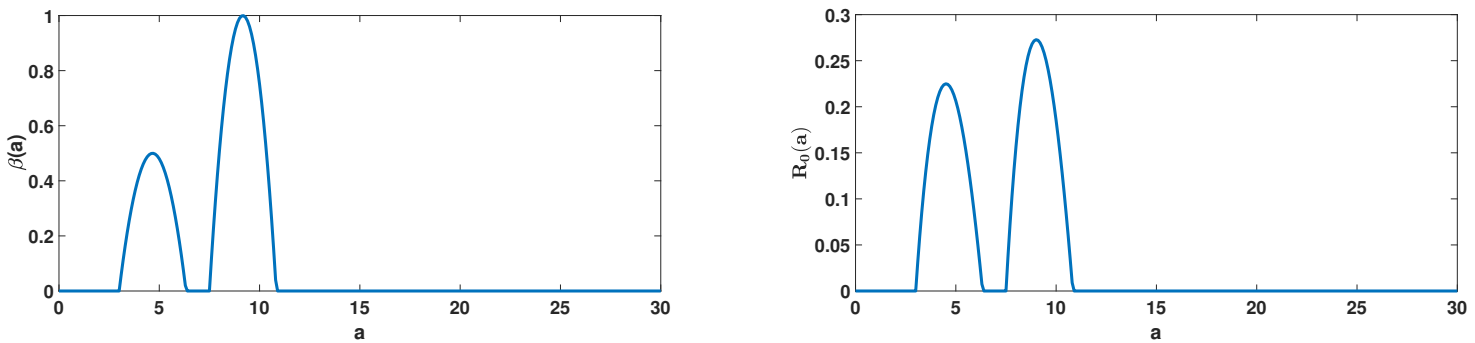


Figure: *On the left-hand side, we plot the function* $a \to \beta(a)$. *On the right-hand side, we plot the function* $a \to R_0(a) = \tau_0 \times S_0 \times \beta(a) \times e^{-\nu a}$.

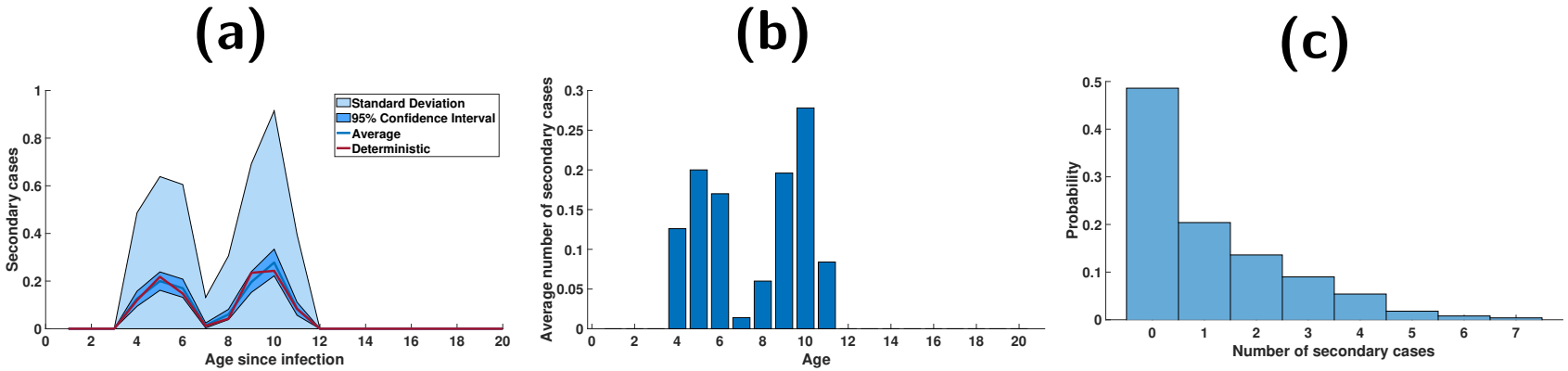


Figure: *In these figures, we present sets of 500 samples of secondary cases produced by a single infected individual in a population of $S = 10^7$ susceptible hosts. Theses samples are produced by using the IBM.* ***(a)*** *Statistical summary: the blue curve represents the average number of cases at age of infection a; the dark blue area is the 95% confidence interval of this average obtained by fitting a Gaussian distribution to the data; the light blue area corresponds to the standard deviation; the orange curve is the deterministic daily basic reproductive number at age a.* ***(b)*** *Bar graph of the average number of secondary cases as a function of the age since infection.* ***(c)*** *Histogram of the total number of secondary cases produced during the whole infection. This estimates the probability of a single infected to generate n secondary cases (with n in the abscissa).*

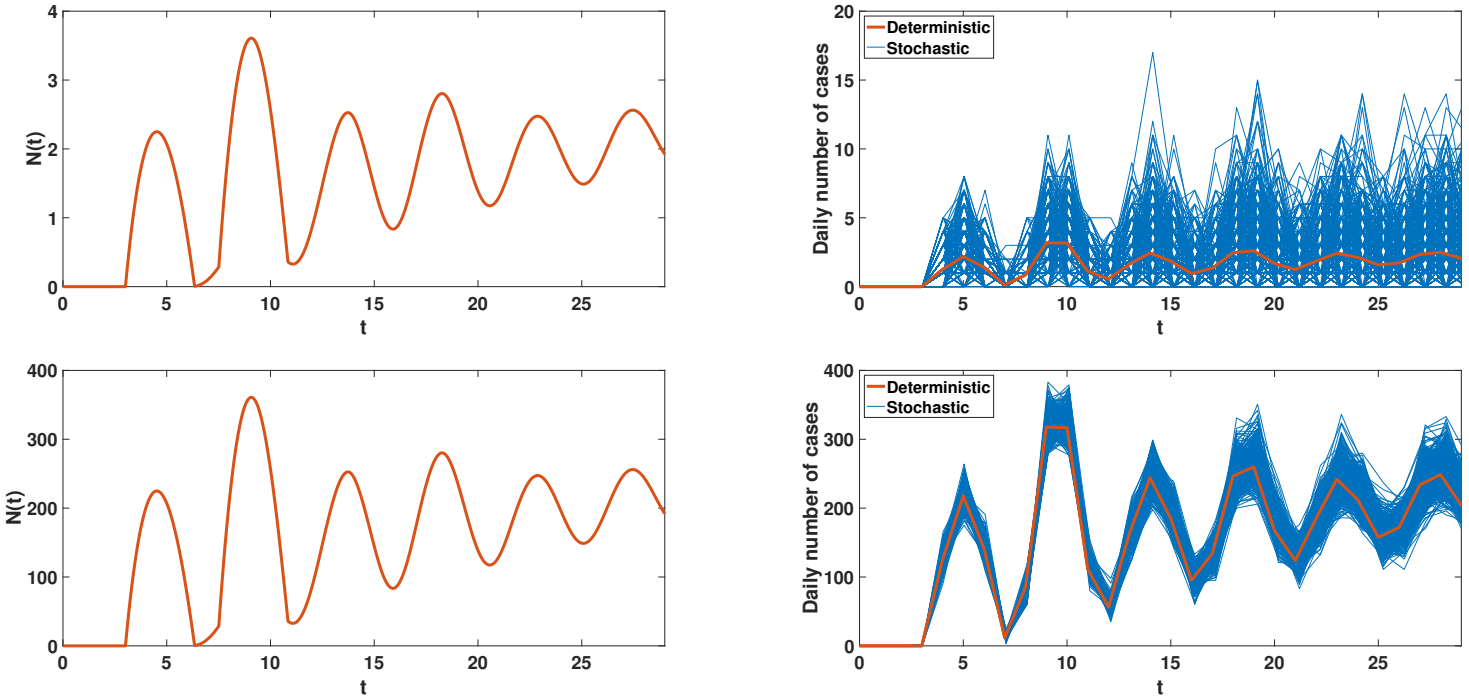


Figure: *On the left-hand side, we plot the function* $t \to N(t)$ *solution of the continuous Volterra integral equation. On the right-hand side, we plot the function* $t \to \int_{t-1}^{t} N(s)ds$ *(for* $t = 1, 2, \ldots$*) which corresponds to the daily number of cases obtained by solving our Volterra equation, and we compare it with the daily number of cases obtained from* 500 *runs of the IBM. The top two figures correspond to* $I_0 = 10$*, and the bottom two figures to* $I_0 = 1\,000$*.*

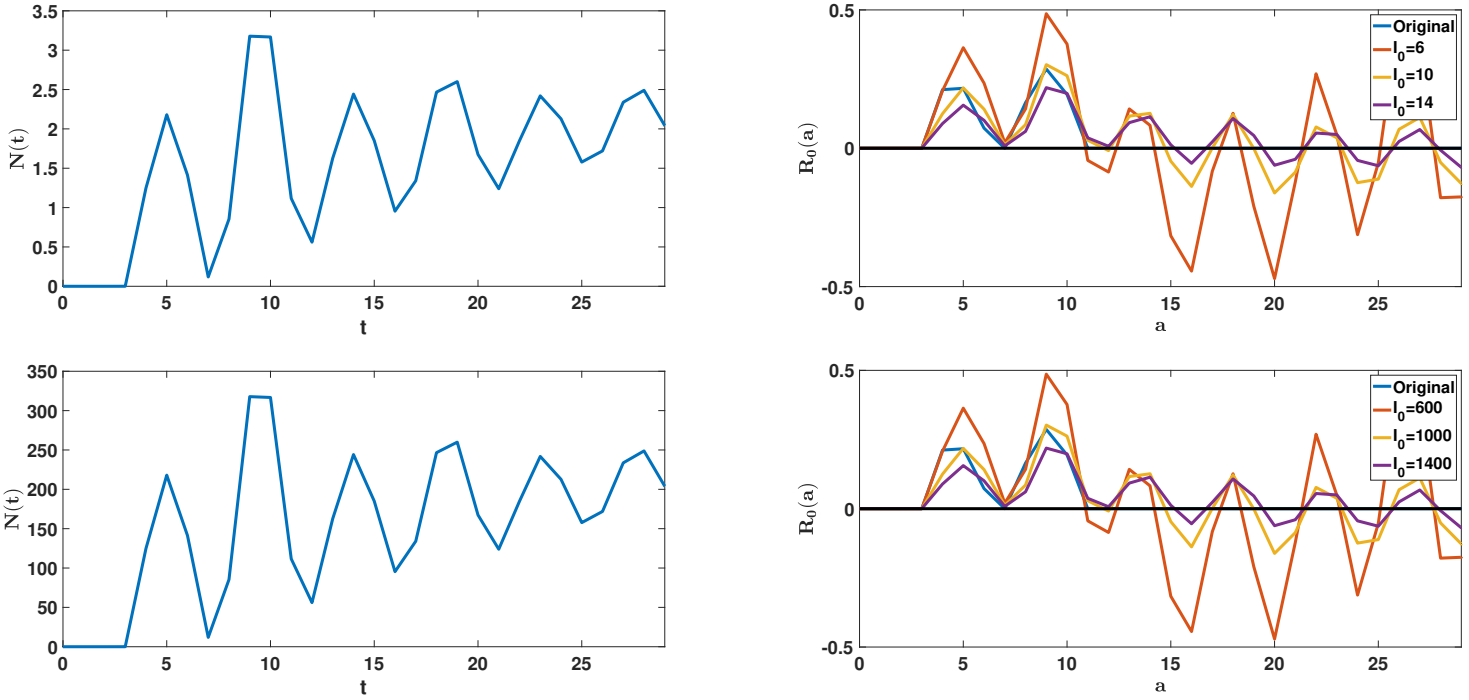


Figure: *On the left hand side, we plot the daily number of* $t \to \int_{t-1}^{t} N(s)ds$ *(for* $t = 0, 1, 2, \ldots$*) by using the continuous Volterra integral equation for* $N(t)$ *with* $I_0 = 10$ *(top) and* $I_0 = 1\,000$ *(bottom). On the right-hand side, we apply formula* (44) *to the flow of new infected obtained from the deterministic model. In the top two figures we vary* $I_0 = 6, 10, 14$*. In the bottom two figures we vary* $I_0 = 600, 1000, 1400$*. In both cases, the yellow curve gives the best visual fit, and the* $R_0(a)$ *becomes negative whenever* I_0 *becomes too small.*

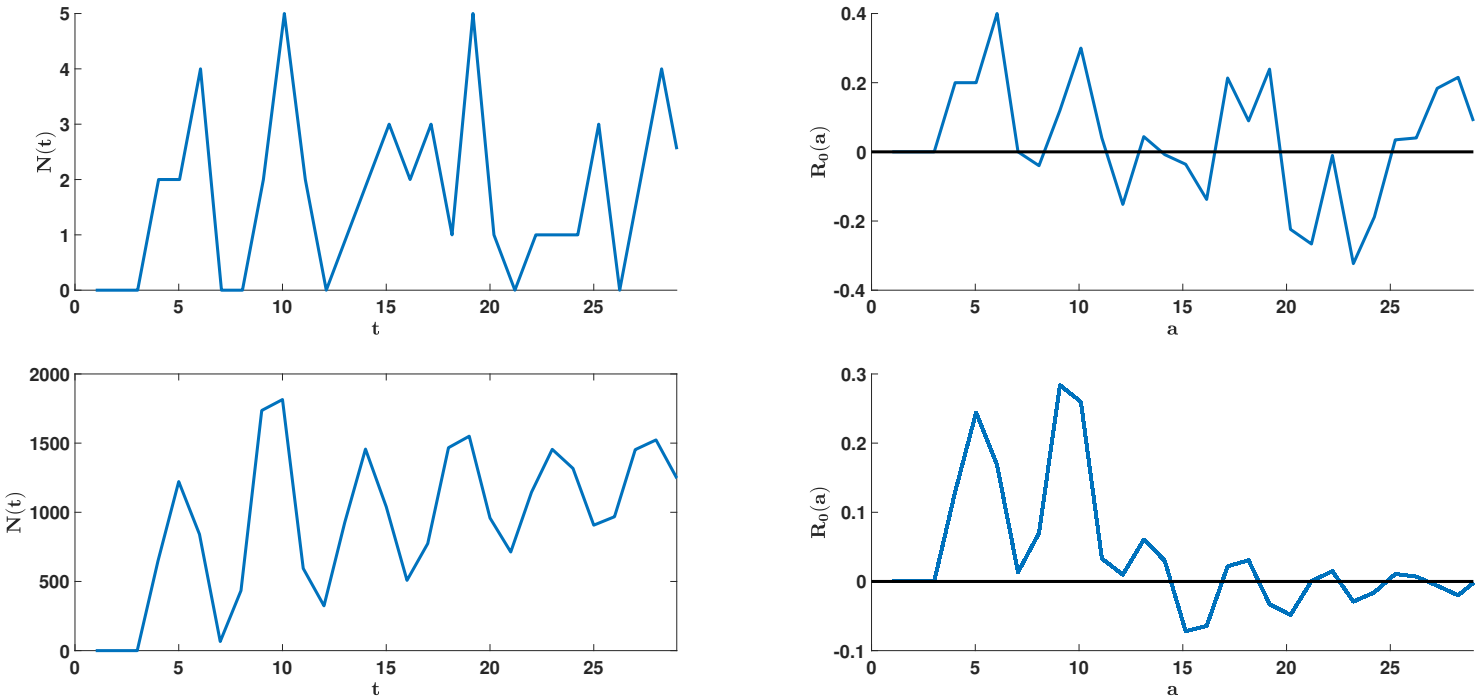


Figure: *On the left-hand side, we plot the daily number of cases* $t \to N(t)$ *(for* $t = 0, 1, 2, \ldots$*) obtained on the top from a single run of the IBM, and the bottom by summing the daily number of cases for* 500 *IMB runs. On the right-hand side, we apply formula the continuous Volterra integral equation for* $R_0(t)$ *(with* $I_0 = 10$*) to the daily number of cases obtained from the IBM. The top two figures correspond to* $I_0 = 10$*, and the bottom two figures to* $I_0 = 500 \times 10$*.*

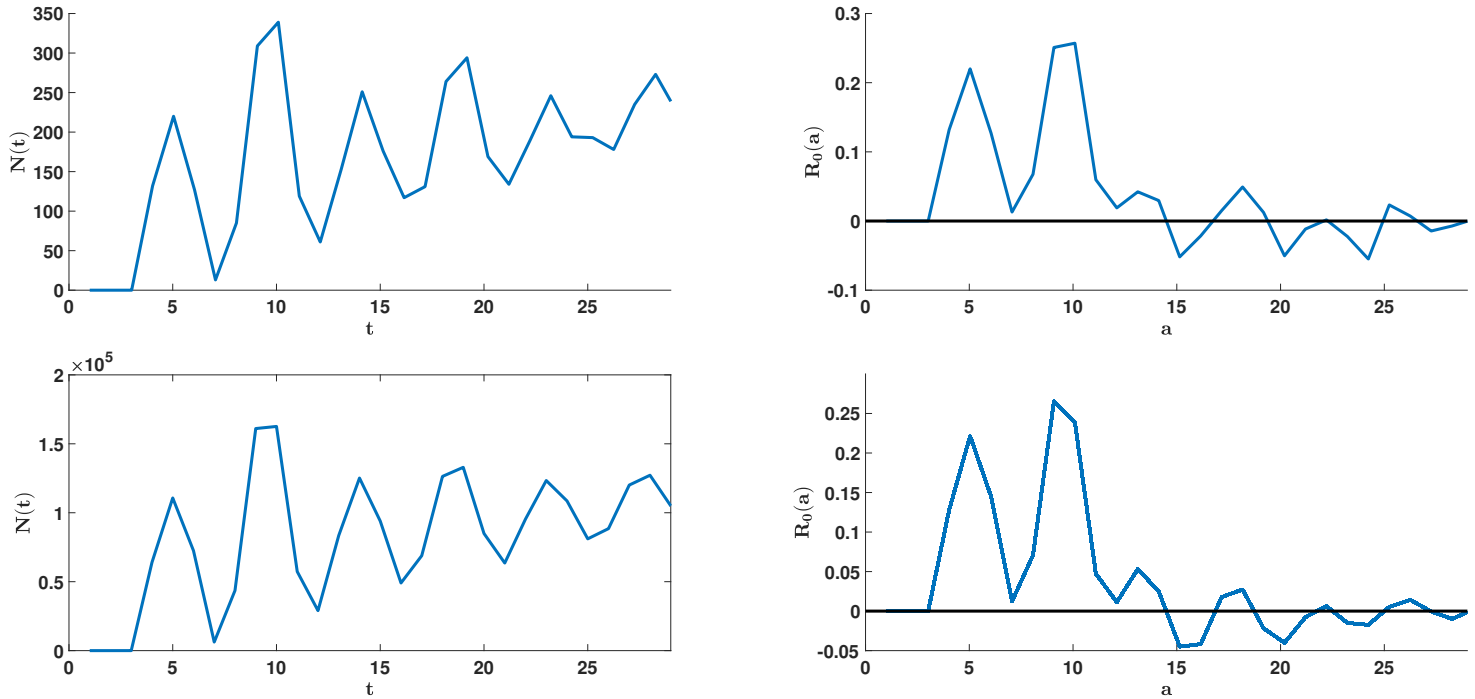


Figure: *On the left-hand side, we plot the daily number of cases* $t \to N(t)$ *(for* $t = 0, 1, 2, \ldots$*) obtained on the top from a single run of the IBM, and the bottom by summing the daily number of cases for* 500 *IMB runs. On the right-hand side, we apply formula the continuous Volterra integral equation for* $R_0(t)$ *(with* $I_0 = 1\,000$*) to the daily number of cases obtained from the IBM. The top two figures correspond to* $I_0 = 1\,000$*, and the bottom two figures to* $I_0 = 500 \times 1\,000$*.*

Application to SARS-CoV-1

(a) **(b)** **(c)**

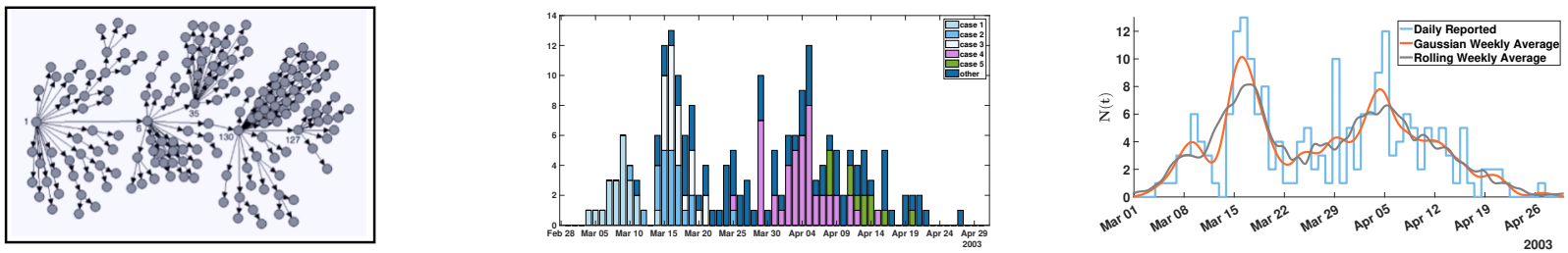


Figure: **(a)** *We plot the contact network of the five super spreader cases in the SARS epidemic in Singapore in 2003 (CDC report). The super spreaders are patient 1, patient 6, patient 35, patient 130 and patient 127.* **(b)** *Daily reported cases from Singapore for the epidemic of SARS in 2003. Case 1 generated 21 cases and 3 suspected cases, case 2 generated 23 cases and 5 suspected cases, case 3 generated 23 cases and 18 suspected cases, case 4 generated 40 cases and 22 suspected cases, case 5 generated 15 cases and 0 suspected cases (CDC report). The cases 1,2,3,4,5 correspond respectively to the patients 1, 6, 35, 130 and 127.* **(c)** *Regularizations of the daily cases data from the SARS-CoV-1 outbreak in Singapore (CDC report). The blue curve corresponds to a step function, the orange curve to a Gaussian weekly average, and the gray curve to a rolling weekly average. The applications in the next Figure are done with the "Rolling Weekly" regularization.*

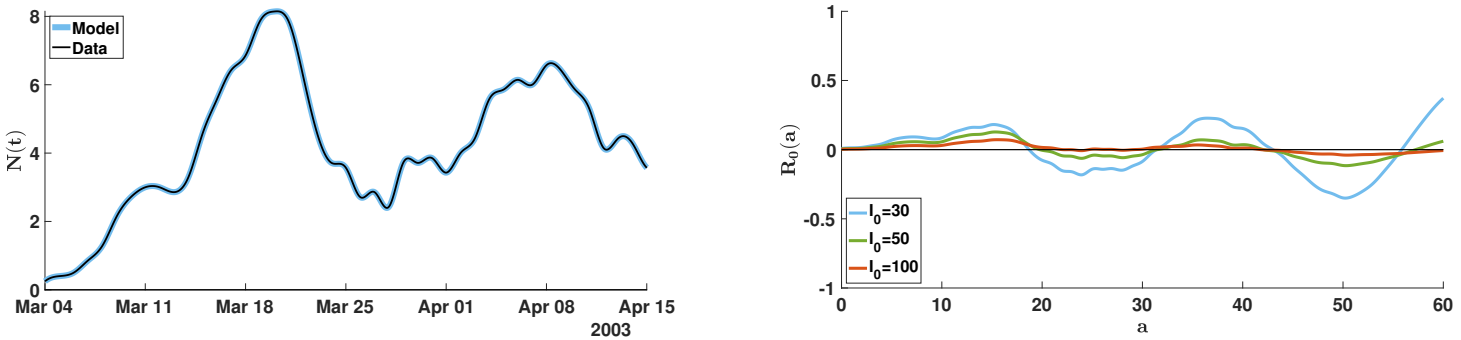


Figure: *Left: Regularized data of the SARS-CoV-1 outbreak in Singapore in 2003 (CDC report) (black line) and the numerical solution of the Volterra integral equation with* $I_0 = 30$ *and* $R_0(a)$ *computed by using the single cohort model. The solutions* $N(t)$ *of the model with* $I_0 = 50$ *and* $I_0 = 100$ *are exactly the same than the data when we use the corresponding* $R_0(a)$*, therefore they are not represented here. Right: numerical solution of the* $R_0(a)$ *function computed by using the continuous model Volterra integral equation with* $I_0 = 30$*,* $I_0 = 50$ *and* $I_0 = 100$*.*

References

- **J. Demongeot and P. Magal**, (2022) Spectral method in epidemic time series, *Biology* 2022, **11(12)**, 1825.

- **J. Demongeot, Q. Griette, Y. Maday, P. Magal**, (2023) A Kermack-McKendrick model with age of infection starting from a single or multiple cohorts of infected patients, *Proceedings of the Royal Society A* 479: 20220381.

Thank you for listening