

Supplementary material to

A Kermack-McKendrick model with age of infection starting from a single or multiple cohorts of infected patients

JACQUES DEMONGEOT^(a), QUENTIN GRIETTE^{(b),*}, YVON MADAY^(c), 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.

^(c) *Sorbonne Université and Université de Paris, CNRS, Laboratoire Jacques-Louis Lions (LJLL), F-75005 Paris, France*

Institut Universitaire de France, 75005 Paris, France

June 16, 2022

1 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 $I_0 \times \kappa \times e^{-\kappa a}$ we obtain

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

In order to derive the Kermack-McKendrick model with Dirac mass initial distribution as limit, we first need the following result.

Lemma 1.1. *Let Assumption 2.1 be satisfied, and assume in addition that $a \rightarrow \beta(a)$ is continuous. Then we have*

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

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

$$\lim_{\kappa \rightarrow +\infty} \sup_{t \geq t_0} |\Lambda_\kappa(t) - I_0 e^{-\nu(t-t_0)} \beta(t-t_0)| = 0.$$

Proof. Let $\varepsilon > 0$. We observe that

$$\begin{aligned} \Lambda_\kappa(t) - I_0 e^{-\nu(t-t_0)} \beta(t-t_0) &= I_0 e^{-\nu(t-t_0)} \int_0^{+\infty} [\beta(a + (t-t_0)) - \beta(t-t_0)] \kappa e^{-\kappa a} da, \\ &= I_0 e^{-\nu(t-t_0)} \int_0^\eta [\beta(a + (t-t_0)) - \beta(t-t_0)] \kappa e^{-\kappa a} da, \\ &\quad + I_0 e^{-\nu(t-t_0)} \int_\eta^{+\infty} [\beta(a + (t-t_0)) - \beta(t-t_0)] \kappa e^{-\kappa a} da. \end{aligned}$$

*Q.G. acknowledges support from ANR via the project Indyana under grant agreement ANR-21-CE40-0008.

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

Let $t_1 > t_0$ be such that

$$I_0 e^{-\nu(t-t_0)} \sup_{a \geq 0} \beta(a) \leq \frac{\varepsilon}{4}, \forall t \geq t_1.$$

Let $\eta > 0$ be such that

$$a \leq \eta \Rightarrow |\beta(a+t) - \beta(t)| \leq \frac{\varepsilon}{2}, \forall t \in [t_0, t_1].$$

Then we have

$$|I_0 e^{-\nu(t-t_0)} \int_0^\eta [\beta(a+(t-t_0)) - \beta(t-t_0)] \kappa e^{-\kappa a} da| \leq \begin{cases} I_0 e^{-\nu(t-t_0)} 2 \sup_{a \geq 0} \beta(a), & \text{if } t \geq t_1, \\ I_0 e^{-\nu(t-t_0)} \int_0^\eta \frac{\varepsilon}{2} \kappa e^{-\kappa a} da, & \text{if } t \in [t_0, t_1], \end{cases}$$

therefore

$$|\Lambda_\kappa(t) - I_0 e^{-\nu(t-t_0)} \beta(t-t_0)| \leq \frac{\varepsilon}{2} + |I_0 e^{-\nu(t-t_0)} \int_\eta^{+\infty} [\beta(a+(t-t_0)) - \beta(t-t_0)] \kappa e^{-\kappa a} da|.$$

The result from the fact that

$$|I_0 e^{-\nu(t-t_0)} \int_\eta^{+\infty} [\beta(a+(t-t_0)) - \beta(t-t_0)] \kappa e^{-\kappa a} da| \leq 2 I_0 \sup_{a \geq 0} \beta(a) \int_\eta^{+\infty} \kappa e^{-\kappa a} da,$$

hence

$$|I_0 e^{-\nu(t-t_0)} \int_\eta^{+\infty} [\beta(a+(t-t_0)) - \beta(t-t_0)] \kappa e^{-\kappa a} da| \leq 2 I_0 \sup_{a \geq 0} \beta(a) e^{-\kappa \eta} \rightarrow 0, \text{ as } \kappa \rightarrow \infty. \quad \square$$

2 Basic reproduction number

In this section, we assume that the transmission $t \rightarrow \tau(t)$ is constant equal to τ , and $t \rightarrow S(t)$ is constant equal to S_0 .

Define the **daily reproduction numbers**

$$R_0(a) = \tau \times S_0 \times \Gamma(a) = \tau \times S_0 \times \beta(a) \times e^{-\nu a}, \forall a \geq 0. \quad (2.1)$$

Assuming that the number of susceptible individuals $t \rightarrow S(t)$ is constant and equal S_0 in the N -equation, then we obtain

$$N(t) = I_0 \times R_0(t-t_0) + \int_0^{t-t_0} R_0(a) N(t-a) da, \quad \forall t \geq t_0. \quad (2.2)$$

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

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

Replacing the notation s by t , and define

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

the equation (2.2) becomes

$$N_{t_0}(t) = \left[I_0 \times R_0(t) + \int_0^t R_0(a) N_{t_0}(t-a) da \right], \forall t \geq 0. \quad (2.3)$$

By replacing $N_{t_0}(t)$ by the right hand side of (2.3) in the integral term of (2.3) we obtain

$$N_{t_0}(t) = I_0 \times R_0(t) + \int_0^t R_0(a) I_0 \times R_0(t-a) da + \int_0^t R_0(a_0) \int_0^{t-a_0} R_0(a_1) N_{t_0}(t-a_0-a_1) da_1 da_0$$

and by induction

$$N_{t_0}(t) = I_0 R_0(t) + I_0 (R_0 * R_0)(t) + I_0 (R_0 * R_0 * R_0)(t) + \dots + I_0 \underbrace{(R_0 * R_0 * \dots * R_0)}_{n \text{ times}}(t) + \dots$$

where the convolution is defined by

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

We define

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

and for each integer $n \geq 3$,

$$(U^{*(n)})(t) = (U^{*(n-1)} * U)(t) = (U * U^{*(n-1)})(t) = \underbrace{(U * U * \dots * U)}_{n \text{ times}}(t).$$

We can interpret the N -equation (2.2) concretely as follows

$$\begin{aligned} N_{t_0}(t) = & \underbrace{I_0 R_0(t)}_{\text{Flow of infected produced at time } t \text{ by the first } I_0 \text{ infected individuals}} \\ & + \underbrace{I_0 (R_0^{*(2)})(t)}_{\text{Flow of infected produced at time } t \text{ by the second generation of infected individuals}} \\ & + \underbrace{I_0 (R_0^{*(3)})(t)}_{\text{Flow of infected produced at time } t \text{ by the third generation of infected individuals}} \\ & + \vdots \\ & + \underbrace{I_0 (R_0^{*(n)})(t)}_{\text{Flow of infected produced at time } t \text{ by the } n^{\text{th}} \text{ generation of infected individuals}} \\ & + \vdots \end{aligned}$$

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 scale 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 .

Proposition 2.1. *The total number of cases produced by the n^{th} generation of infected resulting from a single infected patient is*

$$\int_0^{\infty} \left(R_0^{*(n)} \right) (t) dt = (R_0)^n.$$

Proof. By using Fubini's theorem we have

$$\int_0^{\infty} (R_0 * V)(t) dt = \int_0^{\infty} \int_0^t R_0(t-a)V(a) da dt = \int_0^{\infty} \int_a^{+\infty} R_0(t-a) dt V(a) da$$

and by making the change of variable $l = t - a$ we obtain

$$\int_0^{\infty} (R_0 * V)(t) dt = \int_0^{+\infty} R_0(l) dl \times \int_0^{+\infty} V(a) da.$$

Replacing $V(t)$ by $(R_0^{*(n-1)})(t)$ in the above equation we obtain

$$\int_0^{\infty} \left(R_0^{*(n)} \right) (t) dt = \int_0^{+\infty} R_0(l) dl \int_0^{\infty} \left(R_0^{*(n-1)} \right) (t) dt,$$

and the result follows by induction. □

3 Some explicit examples of $a \rightarrow R_0(a)$

We assume that the transmission $t \rightarrow \tau(t)$ is constant equal to τ , and $t \rightarrow S(t)$ is constant equal to S_0 . Then

$$R_0(a) = \tau \times S_0 \times \Gamma(t - t_0),$$

and by setting $a = t - t_0$, we obtain

$$R_0(a) = \frac{N(a + t_0)}{I_0} - \frac{1}{I_0} \int_0^a R_0(\sigma) N(a - \sigma + t_0) d\sigma, \forall a \geq 0. \quad (3.1)$$

3.1 Exponential decay

Assume that

$$N(t) = \chi_1 I_0 e^{-\chi_2(t-t_0)}.$$

Then we obtain

$$R_0(a) = \chi_1 e^{-\chi_2 a} - \int_0^a R_0(\sigma) \chi_1 e^{-\chi_2(a-\sigma)} d\sigma, \forall a \geq 0.$$

Therefore

$$R_0'(a) = -(\chi_1 + \chi_2) R_0(a), \forall a \geq 0, \text{ and } R_0(0) = \chi_1,$$

and we obtain

$$R_0(a) = \chi_1 e^{-(\chi_1 + \chi_2)a}, \forall a \geq 0. \quad (3.2)$$

3.2 Exponential decay with latency

Let $t_1 > t_0$. Assume that

$$N(t) = \begin{cases} I_0 \chi_1 e^{-\chi_2(t-t_0)}, & \text{if } t \geq t_1, \\ 0, & \text{if } t_0 \leq t < t_1. \end{cases}$$

Then we obtain

$$R_0(a) = 0, \forall a \in [0, t_1 - t_0],$$

and

$$R_0(a) = \chi_1 e^{-\chi_2 a} - \int_{t_1 - t_0}^a R_0(\sigma) \chi_1 e^{-\chi_2(a-\sigma)} d\sigma, \forall t \geq t_1.$$

By setting $a_1 = t_1 - t_0$, we obtain

$$\begin{aligned} R_0(a) &= 0, \forall a \in [0, a_1], \\ R_0(a) &= \chi_1 e^{-\chi_2 a} - \int_{a_1}^a R_0(\sigma) \chi_1 e^{-\chi_2(a-\sigma)} d\sigma, \forall a \geq a_1, \end{aligned}$$

hence by setting $a = \hat{a} + a_1$,

$$R_0(\hat{a} + a_1) = \chi_1 e^{-\chi_2(\hat{a}+a_1)} - \int_{a_1}^{\hat{a}+a_1} \chi_1 e^{-\chi_2(\hat{a}+a_1-\sigma)} R_0(\sigma) d\sigma, \forall \hat{a} \geq 0,$$

and by setting $\hat{a} = a - a_1$, and $\hat{R}_0(\hat{a}) = R_0(\hat{a} + a_1)$, we obtain

$$\hat{R}_0(\hat{a}) = C_1 e^{-\chi_2 \hat{a}} - \int_0^{\hat{a}} \chi_1 e^{-\chi_2(\hat{a}-\sigma)} \hat{R}_0(\sigma) d\sigma, \forall \hat{a} \geq 0,$$

where

$$C_1 = \chi_1 e^{-\chi_2 a_1}.$$

Thus

$$\hat{R}_0(\hat{a}) = C_1 e^{-(\chi_1 + \chi_2)\hat{a}}, \forall \hat{a} \geq 0.$$

We conclude that

$$R_0(a) = \begin{cases} 0, & \text{if } 0 \leq a \leq a_1, \\ \chi_3 e^{-(\chi_1 + \chi_2)a}, & \text{if } a > a_1, \end{cases} \quad (3.3)$$

where

$$\chi_3 = \chi_1 e^{\chi_1 a_1}, \text{ and } a_1 = t_1 - t_0.$$

4 Numerical simulations

4.1 Comparison of deterministic and stochastic simulations

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

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

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.

In the following, we use the numerical scheme described in section 6 to run the simulation of the Volterra integral equation. We use the Individual Based Model (IBM) described in section 7 to run the stochastic simulations of the model. In the following, we illustrate the convergence of the IBM to the deterministic model whenever I_0 increases.

We assume that the probability to be infectious is a shifted gamma like distribution. That is

$$\beta(a) = \beta_0 (a - a_0)^+ e^{-\beta_1 (a - a_0)^+}, \quad (4.1)$$

with

$$a_0 = 3 \text{ days}, \beta_0 = e/2 = 1.3591, \text{ and } \beta_1 = 1/2 = 0.5.$$

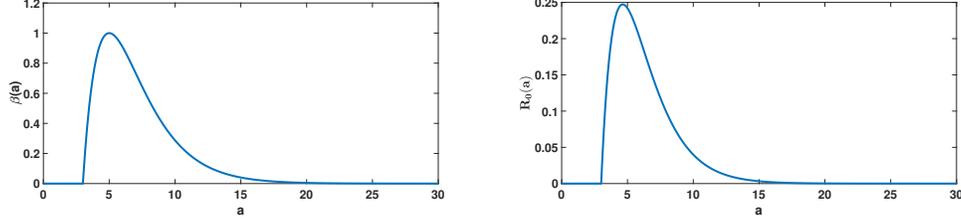


Figure 1: On the left-hand side, we plot the function $a \rightarrow \beta(a)$. On the right-hand side, we plot the function $a \rightarrow R_0(a) = \tau_0 \times S_0 \times \beta(a) \times e^{-\nu a}$.

First generation of secondary cases produced by a single infected: In the Figures 2 we use the IBM to investigate some properties of the clusters obtained from the stochastic simulations. We compare such a stochastic sample with the original $a \rightarrow R_0(a)$.

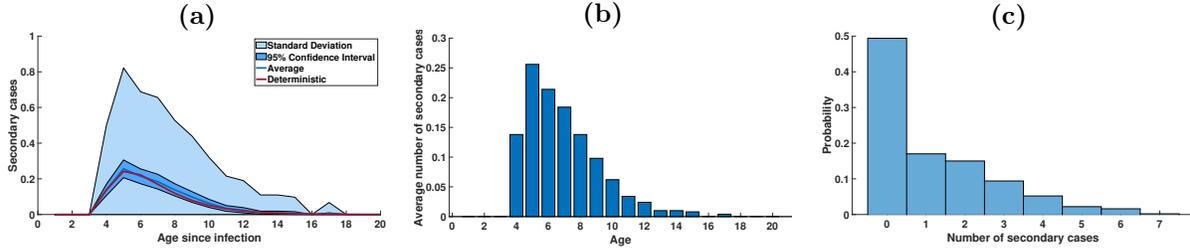


Figure 2: In these figures, we present sets of 500 samples of secondary cases produced by a single infected individual in a population of $S = 10^7$ susceptible hosts. These samples are produced by using the IBM. **(a)** Statistical summary: the blue curve represents the average number of cases at age 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).

Convergence of the IBM to the deterministic model: By comparing Figures 3 and 4 we observe the convergence of the IBM to the deterministic model.

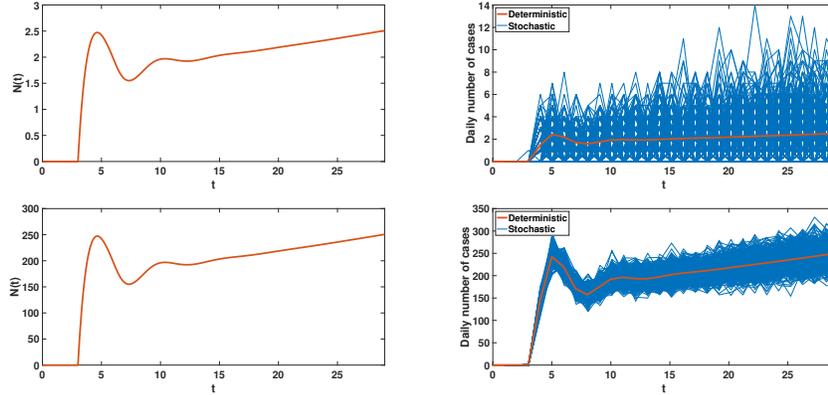


Figure 3: On the left-hand side, we plot the function $t \rightarrow N(t)$ solution of (4.3) with (2.4). On the right-hand side, we plot the function $t \rightarrow \int_{t-1}^t N(s)ds$ (for $t = 1, 2, \dots$) which corresponds to the daily number of cases obtained by solving (4.3) with (2.4), and we 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 $I_0 = 1000$.

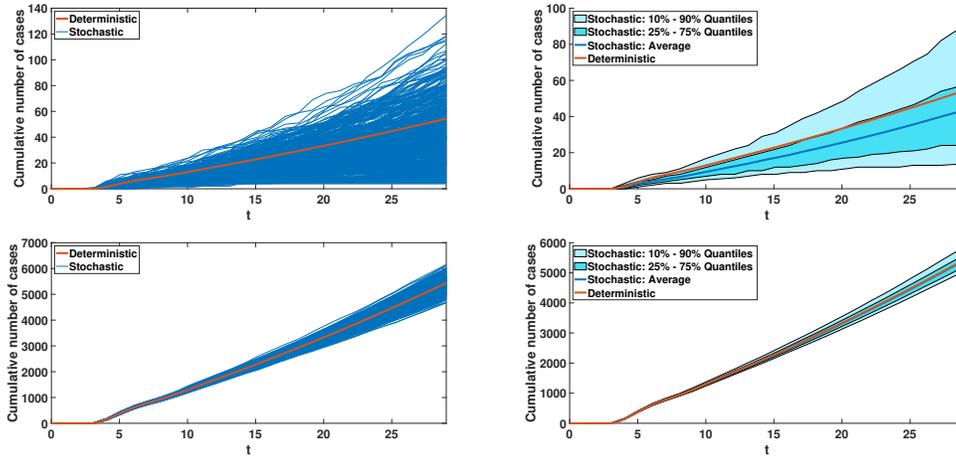


Figure 4: On the left-hand side, we plot the function $t \rightarrow \int_0^t N(s)ds$ (for $t = 0, 1, 2, \dots$) which corresponds to the cumulative number of cases obtained by solving (4.3) with (2.4), and we compare it with the cumulative number of cases obtained from 500 runs of the IBM. On the right-hand side, we plot the average values of the 500 runs obtained from the IBM as well as the quantiles (10%–90% (light blue) and 25%–75% (blue)). The top two figures correspond to $I_0 = 10$, and the bottom two figures to $I_0 = 1000$.

In Figure 5, we focus on the reconstruction of the daily reproduction number from deterministic simulations. In Figure 5, we observe the effect of the day-by-day discretization (which corresponds to the daily reported data). In Figures 6-7, we focus on the reconstruction of the daily reproduction number from stochastic simulations. In Figures 6-7, we observe the stochastic effect of the IBM.

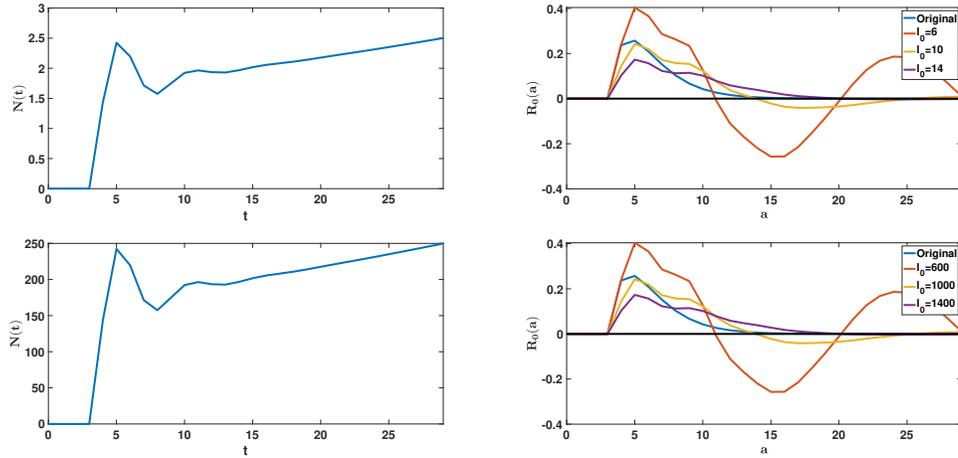


Figure 5: On the left hand side, we plot the daily number of $t \rightarrow \int_{t-1}^t N(s)ds$ (for $t = 0, 1, 2, \dots$) by using the continuous model (1.1) for $I_0 = 10$ (top) and $I_0 = 1000$ (bottom). On the right-hand side, we apply formula (6.5) to the flow of new infected obtained from the 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.

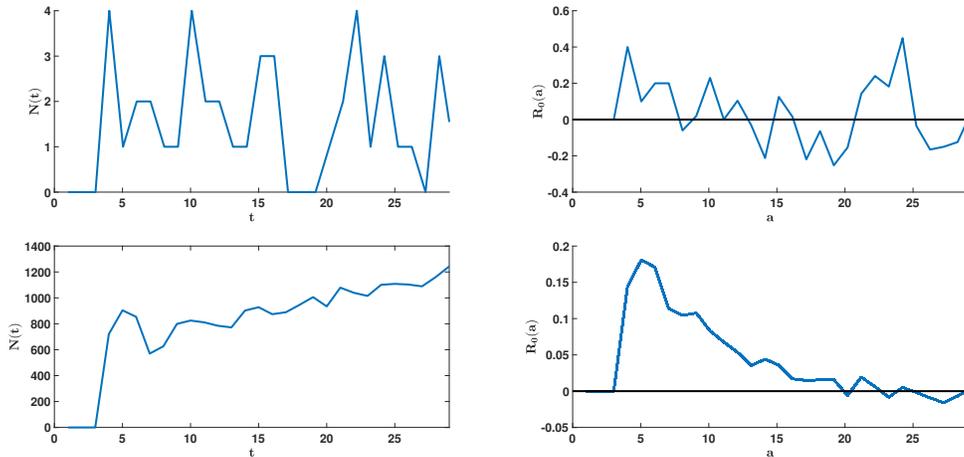


Figure 6: On the left-hand side, we plot the daily number of cases $t \rightarrow N(t)$ (for $t = 0, 1, 2, \dots$) obtained on the top from a single run of the IBM, and the bottom by summing the daily number of cases for 500 runs. On the right-hand side, we apply formula (6.5) (with $I_0 = 10$) to the daily number of cases obtained from the IBM. The top two figures correspond to $I_0 = 10$, and the bottom two figures to $I_0 = 500 \times 10$.

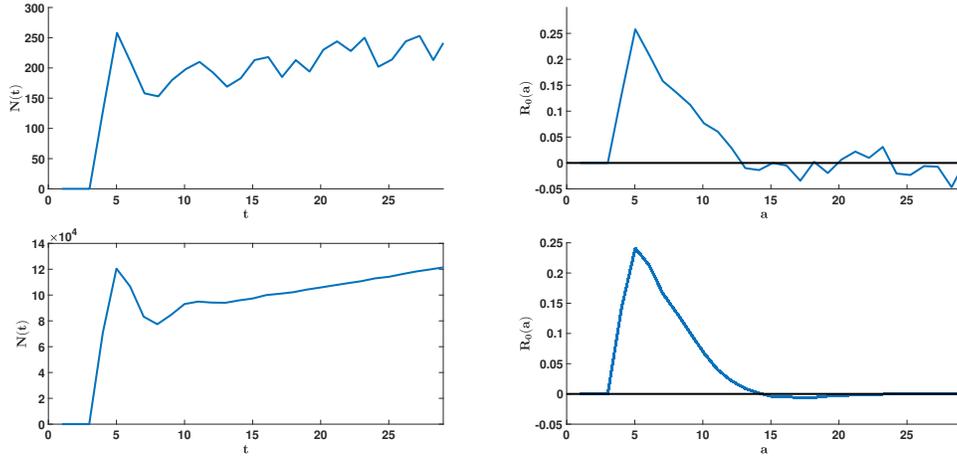


Figure 7: *On the left-hand side, we plot the daily number of cases $t \rightarrow N(t)$ (for $t = 0, 1, 2, \dots$) obtained on the top from a single run of the IBM, and the bottom by summing the daily number of cases for 500 runs. On the right-hand side, we apply formula (6.5) (with $I_0 = 1000$) to the daily number of cases obtained from the IBM. The top two figures correspond to $I_0 = 1000$, and the bottom two figures to $I_0 = 500 \times 1000$.*

The comments on the Figures of the supplementary material are the same as those on the Figures of the main body of the article. The difference is only related to the shape of the distribution of the daily reproduction numbers: it is biphasic in the article, and monophasic in the supplementary material.

Figures 1 to 7 show the influence of the distribution of daily reproduction numbers throughout the period of contagiousness, distribution assumed to be identical for all infected individuals. If it is monophasic, the IBM stochastic approach still allows estimating the daily reproduction numbers with a good precision (Figure 2). This IBM approach coincides with the deterministic one when the size of the simulated sample increases (Figures 3 and 4). The large fluctuations observed in the IBM simulations of the daily reproduction numbers (Figure 3) and of the cumulative cases (Figure 4) are considerably attenuated if we consider the average curves corresponding to different samples of 500 IBM runs.

It can always be noted in Figure 5 that the variations in the daily reproduction numbers of an individual are identical for sets of I_0 equal to 6, 1014 and to 600, 1000, 1400. There is also still a phenomenon of negativity in late daily reproduction numbers, when the duration of the period of contagiousness is high and the initial number of infected is small. This phenomenon is very attenuated in the stochastic model, if we take the average of very many simulations of the IBM model, even in the case where initial number of infected is small (Figures 6 and 7).

Eventually, in the application to the SARS-CoV-1, we see anew the existence of important fluctuations with negative values for the daily reproduction number, especially for late days of the contagiousness period and for small values of I_0 (Figures 9 to 11)

5 Application to SARS-CoV-1

In practice, the Kermack-McKendrick model starting from a Dirac mass means that the epidemic starts from a single patient at time t_0 (whenever $I_0 = 1$) or from a group of I_0 infected patients all with the same age of infection $a = 0$ at time t_0 . This assumption corresponds to the standard conception of a cluster in epidemiology. An example of such a cluster is obtained [33] for the SARS-CoV-1 epidemic in Singapore in 2003. The cluster is represented by a network of contact between individuals in Figure 8.

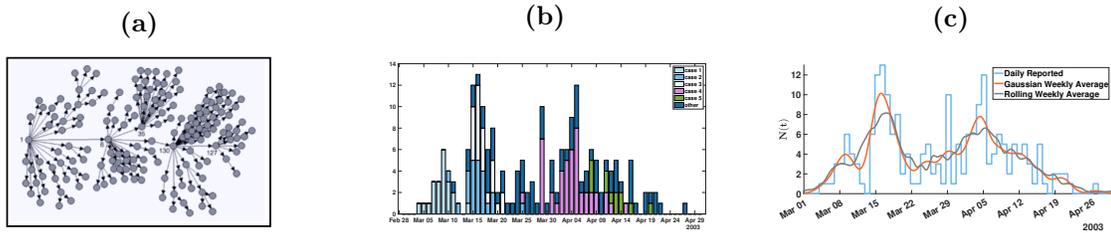


Figure 8: **(a)** We plot the contact network of the five super-spreader cases in the SARS epidemic in Singapore in 2003 [33]. The super-spreaders are patient 1, patient 6, patient 35, patient 130 and patient 127. **(b)** Daily reported cases from Singapore for the epidemic of SARS in 2003. Case 1 generated 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 [33]. 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 [33]. 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 Figure 9 are done with the “Rolling Weekly” regularization.

Figures 8 (a) and (b) present the time series of reported cases by source of infection and date of fever onset and (c) presents three representations of these data in continuous time: as a step function, regularized by Gaussian average and rolling weekly average. In Figure 9 we apply the continuous-time model to the rolling weekly regularization of the data. Similar to the reconstruction of $R_0(a)$ presented in Figures 5-7 and Figures 9-11, the basic reproduction number $R_0(a)$ becomes negative after a given age. Our interpretation is that the data are far from perfect and involves sampling errors and probably a large number of undetected cases. The fact that the transmission rate is subject to variations in time could also explain this negativity. In Figure 10 we apply the discrete model (??) to the original data for different values of I_0 . Finally, in Figure 11, we transform the data by taking advantage of the information on the source of infection given in [33]. We fix an incubation period of 5 days, which corresponds to the average incubation period reported in [33]. Then we shift all secondary cases produced by the six sources identified in the article [33] to the same origin, as if all cases had been produced by the same cluster of six individuals. We present the data on the left-hand side of Figure 11 and apply the method to obtain $R_0(a)$ for parameters $I_0 = 30, I_0 = 50, I_0 = 100$.

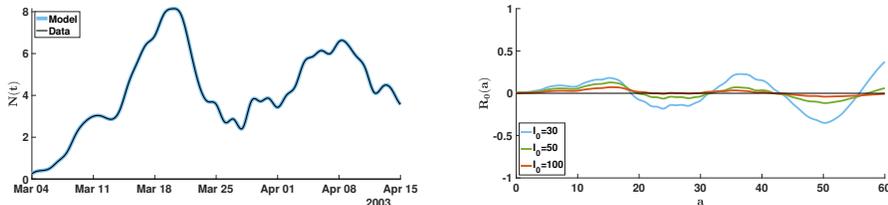


Figure 9: Left: Regularized data of the SARS-CoV-1 outbreak in Singapore in 2003 [33] (black line) and the numerical solution of the model (1.1) with $I_0 = 30$ and $R_0(a)$ computed by (1.6) (blue line). The solutions $N(t)$ of the model (1.1) with $I_0 = 50$ and $I_0 = 100$ are exactly the same when 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 (1.4) with $I_0 = 30, I_0 = 50$ and $I_0 = 100$.

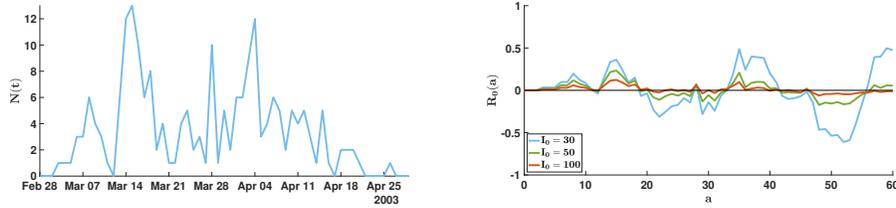


Figure 10: *Left: Original daily reported cases data of the SARS-CoV-1 outbreak in Singapore in 2003 [33] (blue line). Right: Numerical solution of the $R_0(a)$ function computed by using the discrete model (1.6) with $I_0 = 30$, $I_0 = 50$ and $I_0 = 100$.*

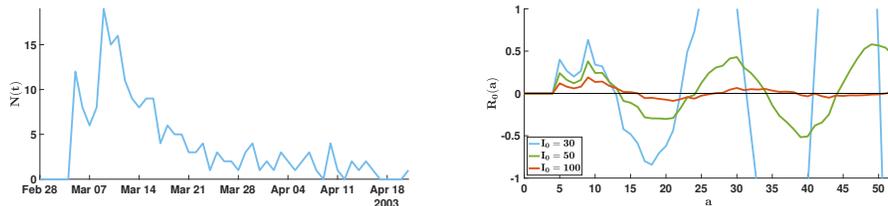


Figure 11: *Left: Daily reported cases of a theoretical cluster based on the data from the SARS-CoV-1 outbreak in Singapore in 2003 [33] (blue line). The secondary cases produced by the six patients identified in [33] are shifted to the same initial infection date of the infector, as if all were produced by the same cluster starting at $t = 0$. We also use an incubation period of 5 days, as indicated in [33]. Right: Numerical solution of the $R_0(a)$ function computed by using the discrete model (1.6) with $I_0 = 30$, $I_0 = 50$ and $I_0 = 100$.*

6 Euler approximation for the Volterra integral equation

We use a numerical scheme for the Volterra integral equation given by

$$N(t) = \tau S(\max(t - \Delta t, t_0)) \left[I_0 \times \Gamma(t - t_0) + \int_0^{\max(t - \Delta t - t_0, 0)} \Gamma(a) N(t - a) da \right], \quad (6.1)$$

with

$$S(t) = S_0 - \int_{t_0}^t N(\sigma) d\sigma, \quad (6.2)$$

and we use the Simpson's rule to compute the integrals.

7 Stochastic simulations: Individual Based Model

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 $\beta(a)$. 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 $1/\nu$.
3. The age of all individuals is updated at fixed intervals of time of size Δt . Simultaneously the life-span of each infected individual 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>.

References

- [1] D. C. Adam, et al. (2020), Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong, *Nature Medicine*, **26.11**, 1714-1719.
- [2] L. Alvarez, M. Colom, J. D. Morel, & J. M. Morel (2021), 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)**.
- [3] L. Alvarez, J-D. Morel, J-M. Morel (2022), Modeling Covid-19 incidence by the renewal equation after removal of administrative bias and noise, *Biology* **11(4)**, 540.
- [4] D. Bernoulli (1760), Essai d'une nouvelle analyse de la mortalité causé par la petite vérole et des avantages de l'inoculation pour la prévenir, *Mém. Math. Phys. Acad. Roy. Sci. Paris*, 1-45.
- [5] M. M. Böhmer, U. Buchholz, V. M. Corman, M. Hoch, K. Katz, D. V. Marosevic, et al. (2020), Investigation of a COVID-19 outbreak in Germany resulting from a single travel-associated primary case: a case series. *The Lancet Infectious Diseases*, **20.8**, 920-928.
- [6] T. C. Chan, and C. C. King (2011), Surveillance and epidemiology of infectious diseases using spatial and temporal clustering methods. In: *Infectious disease informatics and biosurveillance* (207-234). Springer, Boston, MA.
- [7] D. L. Chao, M. E. Halloran, V. J. Obenchain, & Jr, I. M Longini, FluTE, a publicly available stochastic influenza epidemic simulation model. *PLoS computational biology*, **6(1)** (2010), e1000656.
- [8] 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.
- [9] J. Demongeot, Q. Griette, P. Magal, and G. Webb (2022), Modeling vaccine efficacy for COVID-19 outbreak in New York city. *Biology*, **11.3**, 345.
- [10] 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.
- [11] M. R. Desjardins, A. Hohl, and E. M. Delmelle (2020), Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: Detecting and evaluating emerging clusters. *Applied geography* **118**, 102202.
- [12] K. Dietz and J. A. P. Heesterbeek (2002), Daniel Bernoulli's epidemiological model revisited, *Math. Biosci.*, **180**, 1-21.
- [13] A. Ducrot, and P. Magal, A semigroup approach for Volterra integral equation of convolution type (*in preparation*).

- [14] T. Ganyani, C. Kremer, D. Chen, A. Torneri, C. Faes, J. Wallinga, and N. Hens (2020), Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. *Eurosurveillance*, **25.17**, 2000257.
- [15] Q. Griette, J. Demongeot, and P. Magal (2021), A robust phenomenological approach to investigate COVID-19 data for France . *Math. Appl. Sci. Eng.*, 2021.
- [16] 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.
- [17] A. Guttman, L. Ouchchane, X. Li, I. Perthus, J. Gaudart, J. Demongeot, and J. Y. Boire (2013), Performance map of a cluster detection test using extended power. *International Journal of Health Geographics*, **12.1**, 1-10.
- [18] L. Han, P. Shen, J. Yan, Y. Huang, X. Ba, W. Lin, et al. (2021), Exploring the Clinical Characteristics of COVID-19 Clusters Identified Using Factor Analysis of Mixed Data-Based Cluster Analysis. *Frontiers in medicine*, **8**.
- [19] S. Hisada, T. Murayama, K. Tsubouchi, S. Fujita, S. Yada, S. Wakamiya, and E. Aramaki (2020), Surveillance of early stage COVID-19 clusters using search query logs and mobile device-based location information. *Scientific Reports*, **10.1**, 1-8.
- [20] Q. L. Jing, M. J. Liu, Z. B. Zhang, L. Q. Fang, J. Yuan, A. R. Zhang, et al. (2020), Household secondary attack rate of COVID-19 and associated determinants in Guangzhou, China: a retrospective cohort study. *The Lancet Infectious Diseases*, **20.10**, 1141-1150.
- [21] 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.
- [22] C. Kremer, T. Ganyani, D. Chen, A. Torneri, C. Faes, J. Wallinga, and N. Hens. (2020), Authors' response: estimating the generation interval for COVID-19 based on symptom onset data. *Eurosurveillance*, **25.29**, 2001269.
- [23] A. Ladoy, O. Opota, P. N. Carron, I. Guessous, S. Vuilleumier, S. Joost, and G. Greub (2021), Size and duration of COVID-19 clusters go along with a high SARS-CoV-2 viral load: A spatio-temporal investigation in Vaud state, Switzerland. *Science of The Total Environment*, **787**, 147483.
- [24] Z. Liu, P. Magal, O. Seydi, & G. Webb (2020), Understanding unreported cases in the COVID-19 epidemic outbreak in Wuhan, China, and the importance of major public health interventions. *Biology*, **9(3)**, 50.
- [25] H. Nishiura (2007), Time variations in the transmissibility of pandemic influenza in Prussia, Germany, from 1918-19. *Theor. Biol. Med. Model.* **4**, 20.
- [26] H. Nishiura, G. Chowell (2009), "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), 103-121.
- [27] 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.
- [28] R. Pung, C. J. Chiew, B. E. Young, S. Chin, M. I. Chen, H. E. Clapham, et al. (2020), Investigation of three clusters of COVID-19 in Singapore: implications for surveillance and response measures. *The Lancet*, **395.10229**, 1039-1046.
- [29] F. Shams, A. Abbas, W. Khan, U. S. Khan, and R. Nawaz. (2022), A death, infection, and recovery (DIR) model to forecast the COVID-19 spread. *Computer Methods and Programs in Biomedicine Update*, **2**, 100047.

- [30] A. Tariq, Y. Lee, K. Roosa, S. Blumberg, P. Yan, S. Ma, and G. Chowell (2020), Real-time monitoring the transmission potential of COVID-19 in Singapore, March 2020. *BMC medicine*, **18.1**, 1-14.
- [31] M. Worobey, T. Watts, R. McKay, et al. (2016), 1970s and ‘Patient 0’ HIV-1 genomes illuminate early HIV/AIDS history in North America. *Nature* **539**, 98-101. <https://doi.org/10.1038/nature19827>
- [32] S. E. F. Yong, et. al. (2020), Connecting clusters of COVID-19: an epidemiological and serological investigation, *The Lancet Infectious Diseases* **20.7**, 809-815.
- [33] Centers for Disease Control and Prevention (CDC), Severe acute respiratory syndrome Singapore, *MMWR. Morbidity and mortality weekly report* **52.18**, 2005. <https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5218a1.htm>