An extension of Sellke construction and uncertainty quantification for non-Markovian epidemic models

Henri Mermoz Kouye, Clémentine Prieur, Elisabeta Vergu

▶ **To cite this version:**

Henri Mermoz Kouye, Clémentine Prieur, Elisabeta Vergu. An extension of Sellke construction and uncertainty quantification for non-Markovian epidemic models. 2024. hal-04719348

HAL Id: hal-04719348
https://hal.science/hal-04719348v1

Preprint submitted on 3 Oct 2024

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

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

Distributed under a Creative Commons Attribution 4.0 International License

An extension of Sellke construction and uncertainty quantification for non-Markovian epidemic models

Henri Mermoz KOUYE[1,2], Clémentine PRIEUR[2]
and Elisabeta VERGU[1]
[1]Univ. Paris-Saclay, INRAE, MaIAGE, 78350, Jouy-en-Josas, France
[2]Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, LJK, 38000, Grenoble, France

October 3, 2024

Abstract
Several major epidemic events over the past two decades have highlighted the importance of developing and studying non-Markovian compartmental models. Sellke [1983] introduced an ingenious construction for the SIR epidemic process to study the final size of epidemics. In this paper, we extend this construction to the SEI_1I_2RS model. This model is chosen for its compactness, while including parallel infectious stages (I_1 and I_2) and cycles (aka loops) due to reinfection. Our methodology easily generalizes to a general class of stochastic compartmental models in closed populations, including SIR-like models (a series of compartments in one row), SEIAR-like models (parallel compartments), but also models with cycles. Our construction inherits from Sellke construction its ability to handle both Markovian and non-Markovian frameworks. Also, it naturally leads to a representation of the epidemic process under the form of a deterministic function of uncertain parameters (such as epidemic parameters) and variables modeling internal noise. Based on this representation, we propose a global sensitivity analysis of the SEI_1I_2RS model. With our methodology we are able to quantify epistemic uncertainty due to the lack of knowledge on epidemic parameters and statistical uncertainty induced by stochasticity of the model. Finally we provide numerical experiments in both Markovian and non-Markovian frameworks.

Keywords Sellke construction, compartmental models, non-Markovian epidemic process, global sensitivity analysis.

1 Introduction

The COVID-19 pandemic has underlined the importance of mathematical modeling of epidemics. There are a wide variety of mathematical models that can be used

to predict the spread of an epidemic or to guide health decision-making. Models designed to capture the spread of an epidemic in a given population include stochastic compartmental models. These models are related to the seminal model introduced in Kermack and McKendrick [1927] known as the SIR model, where the population is divided into three compartments: susceptible (S), infected (I), and removed (R). The model then follows the change in the proportion of the population belonging to each compartment through time, by reproducing numerically the transition rules from one compartment to the other. Advanced models can be designed with a larger number of compartments to better reflect the characteristics of an outbreak, as well as to match data on which dynamic inference can be performed. Resulting processes can be either Markovian or non-Markovian, depending on modeling framework. Although the Markovian setting is quite restrictive, as it is characterized by memorylessness [Nowzari et al., 2015, Großmann et al., 2021, Sofonea et al., 2021, Saeedian et al., 2017], it has been widely studied in the literature as it is simpler to analyze from a mathematical point of view. Memory in non-Markovian processes is induced by sampling non-exponentially distributed holding times between consecutive state transitions [Van Mieghem and van de Bovenkamp, 2013, Streftaris and Gibson, 2012], or by considering stochastic systems with time delay [see Brett and Galla, 2013, e.g.].

Both Markovian and non-Markovian models are subject to uncertainties. On the one hand, epistemic uncertainty due to a lack of knowledge of the parameters. On the other hand, statistical uncertainty induced by the stochasticity of the model. In compartmental epidemic models, epistemic uncertainty is characterized by imperfect knowledge of the characteristics of the pathogen and the macroscopic behavior of the population, while statistical uncertainty (or internal noise) models the specificity of individuals with respect to infection and behavior. Most risk assessment frameworks for infectious diseases only take into account epistemic uncertainty. However, internal noise has a major impact on the evolution of epidemics that needs to be quantified [Penn et al., 2023]. One way of approaching this issue is to perform a global sensitivity analysis to quantify how the uncertainty in the output of a stochastic model is related to the uncertainty in its parameters, but also to the internal noise. The approach we consider in this paper is based on a representation of the stochastic model under the form of a deterministic function of a set of epidemic parameters and a collection of variables modeling internal noise. This approach has been successfully used in the framework of Markovian models. In Le Maître et al. [2015], the authors used the random time-change representation introduced in Ethier and Kurtz [1986], while the authors in Kouye et al. [2024] used Gillespie algorithms [Gillespie, 1976].

The objective of our work is twofold. We first aim at extending the construction introduced in Sellke [1983] and discussed hereafter to a general class of compartmental models in closed population. We then aim at performing a global sensitivity analysis based on this extended construction, which can handle both Markovian and non-Markovian dynamics.

Sellke [1983] introduced an ingenious construction used to sample the final

size distribution of a SIR epidemic in a finite population. It relies on an individual-based approach and it has the advantage to enable the simulation of a SIR epidemic in both Markovian and non-Markovian frameworks. Since then, it has been studied and extended in many ways. In Reinert [1995], non independent tolerance thresholds of susceptible individuals to infection and non independent infectious periods were considered. Andersson and Britton [2000] studied SIR-multitype epidemic models. In Streftaris and Gibson [2012], the unit-rate exponential distribution of thresholds was changed to a Weibull distribution. House [2014] generalized the construction to the case of finite heterogeneous population, while Di Lauro et al. [2022] focused on the age-dependent SIR model. An extension to the classical SEIR (Susceptible-Exposed-Infectious-Recovered) model was introduced in Britton and Pardoux [2019b]. This extension is straightforward because, on the one hand, the jumps in compartment E follow the same transition mechanism as those in compartment I and, on the other hand, the individuals in compartment E are not infectious. However, in the view of all these works, it appears that so far, the generalizations of Sellke construction mainly focused on SIR-like model. For instance, to the best of our knowledge, none of these works consider models with reinfection and more generally compartmental models that include parallel infectious stages (like SEIAR: Susceptible-Exposed-Infectious-Asymptomatic-Recovered) or such that their corresponding graph includes cycles (or loops) like the classical SIS (Susceptible-Infectious-Susceptible).

In our work, we extend Sellke construction to a more general class of compartmental models in closed populations. Throughout the paper, and for sake of clarity in the presentation, we focus on the SEI_1I_2RS model. However, the methodology easily generalizes to a general class of stochastic compartmental models in closed population, including SIR-like models (a series of compartments in one row), SEIAR-like models (the existence of compartments in parallel), but also models with cycles. The SEI_1I_2RS model was chosen for its compactness, while including multiple parallel infectious stages (I_1 and I_2) and cycles due to reinfection. The construction we propose inherits from the original Sellke construction the ability to cover both Markovian and non-Markovian framework, depending on the choice of probability distributions for holding times between consecutive state transitions. One of the advantages of the construction we propose, and also of the one of Sellke, is that it permits a clear identification of internal noise through the behavior and characteristics of individuals. Using the extended construction, we identify a representation of the epidemic process in the form of a deterministic function of uncertain epidemic parameters and variables modelling internal noise. Based on this representation, we perform global sensitivity analysis of the SEI_1I_2RS model both in Markovian and non-Markovian frameworks.

The paper is organized as follows. In Section 2 we review Sellke original construction (Section 2.1) from which we identify a representation algorithm (Section 2.2) for the SIR model. Then in Section 3 we generalize Sellke construction to a model which includes both loops due to reinfection and two parallel infectious stages, the $SEI1I2RS$ model from which we derive a representation algorithm.

This algorithm is used in Section 4 for simulations. Section 5 presents the methodology and the numerical experiments of global sensitivity analysis based on the representation algorithm. In particular we identify key factors (including epidemic parameters but also variables modeling intrinsic randomness) or key parameter interactions in both Markovian and non-Markovian frameworks. Finally the paper finishes with concluding remarks (Section 6).

2 Sellke original construction, application to the identification of a deterministic representation of stochastic (non)-Markovian SIR model

In Section 2.1 we recall the original Sellke construction from which we identify in Section 2.2 a deterministic representation for the non-Markovian stochastic SIR model.

2.1 Sellke construction

Consider the Susceptible-Infected-Recovered (SIR) model (see Figure 1) and assume that the population is closed and comprises N individuals. We also assume that at each time $t \geq 0$, at most one transition can occur. In the following, we denote by $\{W^\theta(t) = \left(W_S^\theta(t), W_I^\theta(t), W_R^\theta(t)\right)\ , t \geq 0\}$ the jump process counting at each time $t \geq 0$ the number of individuals in each compartment, with $\theta = (\beta, \gamma_I)$, β the transmission rate and $1/\gamma_I$ the mean sojourn time in I.

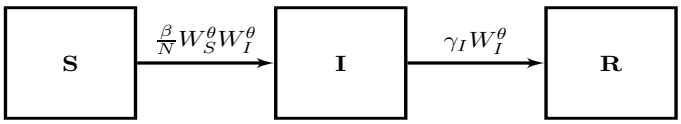

Figure 1: The classical SIR model.

Sellke [1983] proposed an individual-based construction for the SIR model. At the start of the epidemic $(t = 0)$, each individual of the population of size N is labeled $(i = 1, \cdots, N)$ and depending on his initial health status, he is given a set of variables that characterize his behavior towards infection. More precisely, if the individual labeled i is initially susceptible, he is given a threshold $Q_i > 0$ and a sojourn duration $L_i > 0$ in I. If this individual is initially infectious, then he is only given a sojourn duration L_i in compartment I.

Two types of events are responsible for the change of health status: infection or recovery. Occurrence of infection events is linked to a function called the infectious pressure defined as $P(t) = \frac{\beta}{N} \int_0^t W_I^\theta(s)\mathrm{d}s$ so that if the individual labeled i is initially susceptible, he remains susceptible as long as $Q_i > P(t)$ for $t \geq 0$. This function accounts for the pressure exerted by the mass of infectious on susceptible

individuals. At time $\tau_i^S = \inf\{s \geq 0 \mid P(s) \geq Q_i\}$, the individual labeled i gets infected and moves to compartment I (see Figure 2). Individuals with thresholds which remain greater than the value of the infectious pressure escape infection. In Figure 1 below, the labels of individuals have been chosen so that the sequence of tolerance thresholds $(Q_i)_{1 \leq i \leq N}$ is increasing.

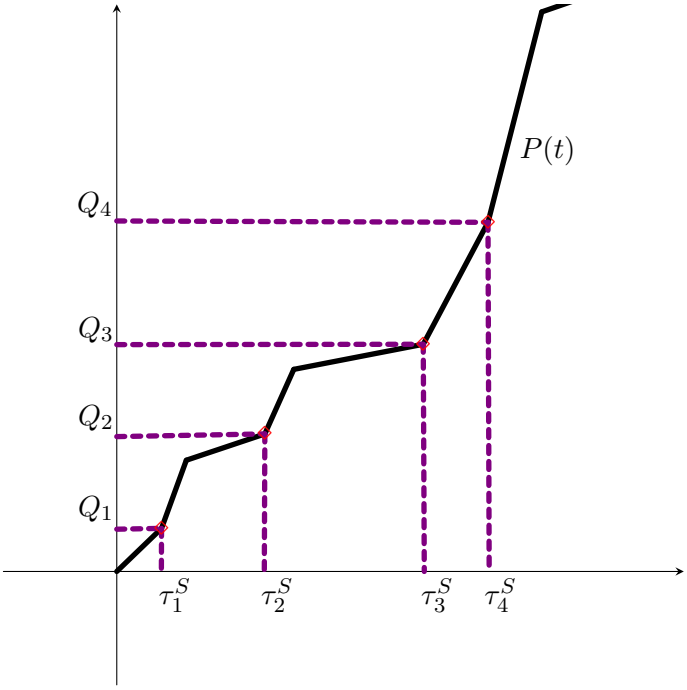

Figure 2: An illustration of the evolution of infectious pressure $t \mapsto P(t)$. Marks $\diamond$ denote infection events. $0 < Q_1 < Q_2 < Q_3 < Q_4$ are tolerance thresholds. $\tau_1^S, \cdots, \tau_4^S$ denote transition times from susceptible to infected associated with thresholds $Q_1, \cdots, Q_4$.

Remark 1. *If the tolerance thresholds $Q_i, i = 1, \ldots, N$ and the sojourn durations $L_i, i = 1, \ldots, N$ are independent from each other and independently distributed under the standard exponential distribution, then the jump process $\{W^\theta(t)\ ,\ t \geq 0\}$ is a continuous-time Markov chain. Moreover, for any time $t \geq 0$, the infection rate is given by $\frac{\beta}{N} W_S^\theta(t) W_I^\theta(t)$ and the recovery rate is given by $\gamma_I W_I^\theta(t)$. This is a classical result which can be found, e.g., in Britton and Pardoux [2019a]. We briefly recall the main lines for the computation of the infection and recovery rates. Let us first focus on the computation of the infection rate. For any $t \geq 0$, the probability that an infection occurs at time $t < s \leq t + \Delta$, conditionally to $W^\theta(t)$, is equal to* $\sum_{i:X_i(t)=S} \mathbb{P}\left(Q_i \leq P(t+\Delta) \mid Q_i > P(t)\right) =$

$\sum_{i:X_i(t)=S} \left(1 - \exp\left(-\frac{\beta}{N} W_I^\theta(t)\Delta\right)\right)$ *as the tolerance thresholds* Q_i *are sampled from the standard exponential distribution. Finally, this last expression is equivalent to* $\frac{\beta}{N} W_S^\theta(t) W_I^\theta(t)\Delta$ *as* $\Delta \to 0$*. Regarding now the computation of the recovery rate, suppose the individual labeled* i *is initially infected (*$t = 0$*) or enters compartment* I *at time* $t > 0$*. Then, the probability that a recovery occurs* $t < s \leq t + \Delta$*, conditionally to* $W^\theta(t)$*, is equal to* $\sum_{i:X_i(t)=I} \mathbb{P}\left(L_i \leq t + \Delta \mid L_i > t\right) = W_I^\theta(t) \times \left(1 - \exp\left(-\gamma_I \Delta\right)\right)$*, which is equivalent to* $\gamma_I W_I^\theta(t)$ *as* $\Delta \to 0$*.*

2.2 A deterministic representation of the SIR model

Recall that the population is supposed closed and that individual initial health statuses, i.e. $X_1(0), \ldots, X_N(0)$, are non-random with values in $\{S, I\}$. Then, it appears that at time $t \geq 0$, the number of susceptible individuals in the population is equal to the number of initially susceptible individuals (i.e. $X_i(0) = S$) whose tolerance threshold remains greater than the value of the infectious pressure at time t, i.e. $P(t)$. Thus

$$W_S^\theta(t) = \sum_{i=1}^{N} \mathbb{1}_{Q_i > P(t), X_i(0)=S} \tag{1}$$

with the convention $Q_i = 0$ if $X_i(0) = I$. By definition, $P(t)$ is deterministic conditionally on $\{W(t)\ ,\ s < t\}$. Thus $W_S^\theta(t)$ is a deterministic function of the sequence of tolerance thresholds conditionally on $\{W(s)\ ,\ s < t\}$.

In order to explicit the number of infected individuals, notice that the compartment I includes at most two types of individuals: either initially infected individuals (infected at $t = 0$) who have not recovered yet, or initially susceptible individuals who are infected but not yet recovered. Thus

$$W_I^\theta(t) = \sum_{i=1}^{N} \mathbb{1}_{L_i > t, X_i(0)=I} + \sum_{i=1}^{N} \mathbb{1}_{\tau_i^S + L_i > t, Q_i \leq P(t), X_i(0)=S}. \tag{2}$$

Recall that τ_i^S is defined as $\tau_i^S = \inf\{s \geq 0 \mid P(s) \geq Q_i\} = P^{-1}(Q_i)$ with P^{-1} the generalized inverse of P. Thus, as P is non decreasing, $\tau_I^S =$ is a deterministic function of Q_i conditionally on $\{W(s)\ ,\ s < t\}$. This yields, together with Equation (2), that $W_I^\theta(t)$ is a deterministic function of the sequence of tolerance thresholds and the sequence of sojourn durations, conditionally on $\{W(s)\ ,\ s < t\}$. Relying on the assumption that the population is closed, it holds that for any $t \geq 0$, $W_S^\theta(t) + W_I^\theta(t) + W_R^\theta(t) = N$. Thus,

$$W_R^\theta(t) = N - W_S^\theta(t) - W_I^\theta(t)\ \forall\, t \geq 0, \tag{3}$$

so that $W_R^\theta(t)$ is a deterministic function of the sequence of tolerance thresholds and the sequence of sojourn durations, conditionally on $\{W(s)\ ,\ s < t\}$. Finally, for $\theta := (\beta, \gamma_I)$ and Z a standardized version of $\widetilde{Z} := \{(Q_1, L_1), \ldots, (Q_N, L_N)\}$,

we deduce from Equations (1), (2) and (3) that there exists a deterministic function f such that, for all $t \geq 0$ and conditionally on $\{W^\theta(s), 0 \leq s < t\}$, $W^\theta(t) := \left(W_S^\theta(t), W_I^\theta(t), W_R^\theta(t)\right) = f(t, \theta, Z)$. Moreover, if θ is sampled from a random vector $\mathbf{X}$ independent of Z, then (f, Z) defines a deterministic representation of $(\mathbf{X}, W^{\mathbf{X}})$ in the sense that:

$$\left(\mathbf{X}, W^{\mathbf{X}}(\cdot)\right) \overset{\mathcal{D}}{=} \left(\mathbf{X}, f(\cdot, \mathbf{X}, Z)\right). \tag{4}$$

Such a representation will be a key stone for performing global sensitivity analysis (see Section 5 for more details).

3 Extending Sellke construction

The construction of Sellke [1983] was originally introduced to study the distribution of the final size of epidemics modeled by the SIR model. A natural question is whether such a construction can be extended to more complex compartmental models. The answer is positive. In this section we focus on the extension of Sellke construction to the SEI_1I_2RS model. This model is interesting as it includes loops (individuals may be reinfected) and a branching compartment, i.e. a compartment which leads to more than one compartment, like compartment E in the SEI_1I_2RS model.

3.1 Description of the SEI1I2RS model

The SEI_1I_2RS model presented in Figure 3 is a compartmental model with health statuses $\{S, E, I_1, I_2, R\}$ and six different types of transitions $\{(S, E), (E, I_1), (E, I_2), (I_1, R), (I_2, R), (R, S)\}$. The different types of transitions between states are described in Table 1.

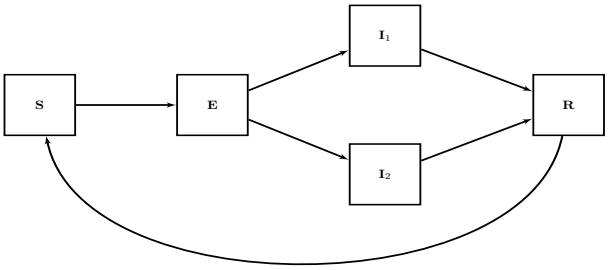

Figure 3: The SEI_1I_2RS model

This model enables to describe the propagation of an epidemic within a population with the following characteristics: existence of an incubation period for the infected individuals, the presence of two categories of infectious individuals, and the absence of permanent immunity (possible reinfection). For instance, it

could be used to model the spread of SARS-CoV-2 by considering individuals in compartment I_1 as symptomatic and those in compartment I_2 as asymptomatic. To each type of transitions we associate a vector $u = (u_1, u_2, u_3, u_4, u_5)$, with $u_i \in \{-1, 0, 1\}$ (see last column of Table 1).

Type of transition	Type	Transition vector
(S, E)	infection	$\mathbf{u}_{(S,E)} := (-1, 1, 0, 0, 0)$
(E, I_1)	first type infection activation	$\mathbf{u}_{(E,I_1)} := (0, -1, 1, 0, 0)$
(E, I_2)	second type infection activation	$\mathbf{u}_{(E,I_2)} := (0, -1, 0, 1, 0)$
(I_1, R)	first type recovery	$\mathbf{u}_{(I_1,R)} := (0, 0, -1, 0, 1)$
(I_2, R)	second type recovery	$\mathbf{u}_{(I_2,R)} := (0, 0, 0, -1, 1)$
(R, S)	reinfection	$\mathbf{u}_{(R,S)} := (1, 0, 0, 0, -1)$

Table 1: Description of the model transitions between states.

The succession over time of these transitions completely defines the states of the epidemic process. The dynamics depend on a certain number of parameters related to the epidemic characteristics. In the current modeling context, the SEI_1I_2RS model depends on six epidemic parameters listed in Table 2.

Parameter Name	Parameter Role
β	transmission rate
$1/\mu_E$	mean sojourn duration in E
p	probability for an exposed individual to move to I_1
$1/\mu_1$	mean sojourn time in I_1
$1/\mu_2$	mean sojourn time in I_2
$1/\delta$	mean sojourn time in R

Table 2: Epidemic parameters.

In the following, $\theta := (\beta, \mu_E, p, \mu_1, \mu_2, \delta)$ is defined as the vector of epidemic parameters. The set of values for θ is denoted by $\mathcal{X} \subseteq \mathbb{R}^6$. We denote by $W(t) := (W_S(t), W_E(t), W_{I_1}(t), W_{I_2}(t), W_R(t))$ so that $W = \{W(t), t \geq 0\}$ represents the epidemic process corresponding to the SEI_1I_2RS model.

3.2 Extension of Sellke construction for the SEI1I2RS model

The construction introduced in Sellke Sellke [1983] for the SIR model cannot be directly applied to the SEI_1I_2RS model as for this last model, there exists a branching compartment (compartment E) and the possibility of reinfection ($R \to S$). The existence of the branching compartment E induces a bifurcation thus, preliminary to any transition from E, one has to active a selection mechanism to choose either $E \to I_1$ or $E \to I_2$. Reinfection introduces loops in the structure of the model. Indeed, it is possible for an individual to exit a compartment and to return back to it after a finite number of transitions. The infection mechanism described in Section 2 has thus to be generalized in a suitable way. This is the purpose of what follows.

Infection mechanism: transitions $S \to E$

As in Section 2, the infection mechanism still depends on a pressure function. There are two groups of infectious individuals which exert pressure over susceptible individuals within the population. We define the pressure function as:

$$P : t \mapsto \frac{\beta}{N} \int_0^t W_I(s)\mathrm{d}s \tag{5}$$

with for any $s \geq 0$, $W_I(s) = W_{I_1}(s) + W_{I_2}(s)$. With this formulation, we assume that individuals in I_1 and I_2 share a common transmission rate β. One may assume instead that each infectious group is characterized by its own transmission rate, say β_1 and β_2. In that case, the pressure function would be given by $t \mapsto \int_0^t (\frac{\beta_1}{N} W_{I_1}(s) + \frac{\beta_2}{N} W_{I_2}(s))\mathrm{d}s$.

We now describe the infection mechanism. If the individual labeled i is susceptible at initial time $t = 0$, he is assigned an initial tolerance pressure $Q_{i,0} > 0$ and he gets infected as soon as $t \geq Q_{i,0}$. Otherwise $Q_{i,0} = 0$. Then, each time the individual labeled i becomes susceptible, he is assigned a new tolerance threshold $Q_{i,j+1} > 0$ and he remains susceptible as long as the excess pressure remains below that threshold. More precisely, if we denote by $\eta_{i,j}^S$ the entrance times of individual labeled i in compartment S, then for $t \geq \eta_{i,j}^S$, the individual labeled i remains susceptible as long as $P(t) - P(\eta_{i,j}^S) < Q_{i,j+1}$. He gets infected at the first time the pressure function reaches and exceeds $P(\eta_{i,j}^S) + Q_{i,j+1}$, i.e. $\tau_{i,j+1}^S = \inf\{t \geq \eta_{i,j}^S \mid P(t) \geq P(\eta_{i,j}^S) + Q_{i,j+1}\}$. In the following we denote by Q_i the sequence $(Q_{i,0}, Q_{i,1}, \ldots)$. We also remark that the pressure function computed at time t, $P(t)$, is deterministic conditionally on $\{W(s)\ ,\ s < t\}$.

Handling transitions from compartments E, I_1, I_2 and R

Transitions from compartments E, I_1, I_2 and R are handled as transition from I in the original Sellke construction. As soon as an individual enters one of

these compartments, he is assigned a sojourn duration. At the end of this period, the individual leaves the compartment. In the following, we denote by $L_i^\alpha = (L_{i,0}^\alpha, L_{i,1}^\alpha, \ldots)$ the succesive sojourn durations of individual labeled i in compartment $\alpha \in \{E, I_1, I_2, R\}$, with $L_{i,0}^\alpha = 0$ except if individual labeled i is in compartment α at initial time $t = 0$. For an individual in compartment E, the transition leads to compartment I_1 with a probability p in $(0, 1)$ and to compartment I_2 with probability $1 - p$. The selection mechanism is modeled with a multinomial probability distribution (binomial in the present setting with only two infectious compartments), denoted by $\mathcal{M}(\{E \to I_1, E \to I_2\}, \mathbf{p} = (p, 1-p))$. Then, each time an individual leaves compartment E, a random variable is sampled from this multinomial distribution to select either $E \to I_1$ or $E \to I_2$. The successive sampling for individual labeled i are denoted by $M_i = (M_{i,1}, \ldots)$.

Retrieving the Markovian case

Similarly to what happens for the construction introduced by Sellke for the SIR model, the transition mechanisms we introduced above lead to a continuous-time Markov chain if one samples tolerance thresholds and sojourn durations adequately. The precise result is stated in Proposition 1 below.

Proposition 1. *Assume that*

- *the sequences Q_i, L_i^α, $\alpha \in \{E, I_1, I_2, R\}$ and M_i, $1 \leq i \leq N$ are mutually independent,*
- *the variables in each of these sequences are independent and identically distributed,*
- *the tolerance thresholds are sampled from the standard exponential distribution,*
- *the sojourn durations in comparment α are sampled from the exponential distribution with mean $1/\lambda_\alpha$, with:*

$$\lambda_\alpha = \begin{cases} \mu_E & \text{if } \alpha = E, \\ \mu_1 & \text{if } \alpha = I_1, \\ \mu_2 & \text{if } \alpha = I_2, \\ \delta & \text{if } \alpha = R. \end{cases}$$

Then, the process $\{W(t)\,,\, t \geq 0\}$ is a homogeneous continuous-time Markov chain with transition rate functions provided in Table 3 below.

Type of transition	Rates
(S, E)	$\frac{\beta}{N} \cdot W_S \cdot (W_{I_1} + W_{I_2})$
(E, I_1)	$p \cdot \mu_E \cdot W_E$
(E, I_2)	$(1-p) \cdot \mu_E \cdot W_E$
(I_1, R)	$\mu_1 \cdot W_{I_1}$
(I_2, R)	$\mu_2 \cdot W_{I_2}$
(R, S)	$\delta \cdot W_R$

Table 3: Transition rates for the Markovian model.

The proof of Proposition 1 is based on classical arguments. It is postponed to the Appendix A.

3.3 A deterministic representation of the SEI1I2RS model

As in the original Sellke construction, assume that the N individuals of the closed population are labeled $i = 1, \ldots, N$. Recall that the health status of the individual labeled i at time $t \geq 0$ is the random variable $X_i(t)$ taking values in $\mathbf{V} = \{S, E, I_1, I_2, R\}$. As already mentioned, such an individual is characterized by a vector of sequences $(Q_i, L_i^E, L_i^I, L_i^R, M_i)$. Moreover these sequences are infinite as, due to the presence of loops in the model, an individual may visit a compartment infinitely many times. Let us now define $\tau_{i,j}^S$, $j \geq 1$, as the jth exit time of individual labeled i from compartment S. We set $\tau_{i,0}^S = 0$ except if individual labeled i is in compartment S at time $t = 0$, in which case we set $\tau_{i,0}^S = \inf\{t > 0 \text{ such that } P(t) \geq Q_{i,0}\}$. For $j \geq 0$, the following recurrence relation holds:

$$\tau_{i,j+1}^S = \inf\{t > \tau_{i,j}^S + \Delta_{i,j+1}^S \text{ such that } P(t) \geq P(\tau_{i,j}^S + \Delta_{i,j+1}^S) + Q_{i,j+1}\} \quad (6)$$

with $\Delta_{i,j+1}^S$ the cumulated sojourn duration of individual number i in compartments other than S before his $(j+1)$th arrival in S. It is given by:

$$\Delta_{i,j+1}^S = \begin{cases} L_{i,j+1}^E + L_{i,j+1}^I + L_{i,j+1}^R \text{ if } X_i(0) = S, \\ L_{i,j}^E + L_{i,j+1}^I + L_{i,j+1}^R \text{ if } X_i(0) = E, \\ L_{i,j}^E + L_{i,j}^I + L_{i,j+1}^R \text{ if } X_i(0) = I, \\ L_{i,j}^E + L_{i,j}^I + L_{i,j}^R \text{ if } X_i(0) = R. \end{cases} \quad (7)$$

Note that $\tau_{i,j}^S + \Delta_{i,j+1}^S$ is the time when the individual labeled i becomes susceptible for the $j+1$-th time. Thus, as long as the infection pressure is below $P(\tau_{i,j}^S + \Delta_{i,j+1}^S) + Q_{i,j+1}$, this individual remains in S. Overall, it turns out that the number

of susceptible individuals can be reformulated as:

$$W_S(t) = \sum_{i=1}^{N} \sum_{j \geq 0} \mathbb{1}_{\tau^S_{i,j} + \Delta^S_{i,j+1} \leq t < \tau^S_{i,j+1}}. \tag{8}$$

Now, by Equation (6) and System (7), and as P is non decreasing with $P(t)$ deterministic conditionally on $\{W(s)\ ,\ s < T\}$, the sequence $\{\tau^S_{i,j}, j \geq 0\}$ is a deterministic function of $\{Q_i, L^E_i, L^I_i, L^R_i\}$ conditionally on $\{W(s)\ ,\ s < t\}$. We then deduce from Equation (8) that $W_S := \{W_S(t), t \geq 0\}$ is a deterministic function of $\{Q_i, L^E_i, L^I_i, L^R_i, i = 1, \ldots, N\}$ conditionally on $\{W(s)\ ,\ s < t\}$.

Similarly, exit times of individuals from other compartments obey explicit recurrence formulas. Indeed, if $\alpha \in \mathbf{V} \setminus \{S\}$, it holds that:

$$\tau^\alpha_{i,j+1} = \tau^\alpha_{i,j} + \Delta^\alpha_{i,j+1} \tag{9}$$

with

$$\Delta^\alpha_{i,j+1} = \begin{cases} \tau^S_{i,j} - \tau^S_{i,j-1} + \sum_{S < \gamma \leq \alpha} \left(L^\gamma_{i,j+1} - L^\gamma_{i,j} \right) \text{ if } X_i(0) = S, \\ \\ \tau^S_{i,j} - \tau^S_{i,j-1} + \sum_{S < \gamma \leq X_i(0)} \left(L^\gamma_{i,j} - L^\gamma_{i,j-1} \right) \\ \qquad + \sum_{X_i(0) < \gamma \leq \alpha} \left(L^\gamma_{i,j+1} - L^\gamma_{i,j} \right) \text{ if } X_i(0) > S \text{ and } \alpha > X_i(0), \\ \\ \tau^S_{i,j+1} - \tau^S_{i,j} + \sum_{S < \gamma \leq \alpha} \left(L^\gamma_{i,j+1} - L^\gamma_{i,j} \right) \\ \text{ if } X_i(0) > S \text{ and } \alpha \leq X_i(0), \end{cases} \tag{10}$$

and with $\tau^\alpha_{i,0} = 0$ except if $X_i(0) = \alpha$ in which case $\tau^\alpha_{i,0} = L^\alpha_{i,0}$. Details leading to (10) are postponed to Appendix B. The quantity $\Delta^\alpha_{i,j+1}$ corresponds to the time spent by the individual labeled i outside the α compartment before his $(j+1)$-th visit to α, added to the time he spends in α during this new stay. Thus, $\tau^\alpha_{i,j} + \Delta^\alpha_{i,j+1} - L^\alpha_{i,j+1}$ corresponds to the time at which this individual enters the compartment α so that for $t \in [\tau^\alpha_{i,j} + \Delta^\alpha_{i,j+1} - L^\alpha_{i,j+1}, \tau^\alpha_{i,j+1})$, $X_i(t) = \alpha$. Thus:

$$W_\alpha(t) = \sum_{i=1}^{N} \sum_{j \geq 0} \mathbb{1}_{\tau^\alpha_{i,j} + \Delta^\alpha_{i,j+1} - L^\alpha_{i,j+1} \leq t < \tau^\alpha_{i,j+1}}. \tag{11}$$

Now, by Equation (9) and System (10), and as P is non decreasing with $P(t)$ deterministic conditionally on $\{W(s)\ ,\ s < t\}$, the sequences $\{\tau^\alpha_{i,j}, j \geq 0\}$ and $\{\Delta^\alpha_{i,j}, j \geq 1\}$ are deterministic functions of $\{Q_i, L^E_i, L^I_i, L^R_i\}$ conditionally on $\{W(s)\ ,\ s < t\}$. We then deduce from Equation (11) that $W_\alpha := \{W_\alpha(t), t \geq 0\}$ is a deterministic function of $\{Q_i, L^E_i, L^I_i, L^R_i, i = 1, \ldots, N\}$ conditionally on $\{W(s)\ ,\ s < t\}$.

The description of the different exit times through recurrence formulas enables to build an algorithm that describes any trajectory of the epidemic process. Such

an algorithm is presented in Algorithm 1 below. As in Table 1, for any $(\alpha, \alpha') \in \mathbf{V} \times \mathbf{V}$, the transition vector $\mathbf{u}_{\alpha,\alpha'}$ is a 5-dimensional vector.

In Algorithm 1, RG_Q denotes the sequence of thresholds $(Q_1, \ldots, Q_N)$ with $Q_i = (Q_{i,0}, Q_{i,1}, \ldots)$. If $X_i(0) = S$, then $(Q_{i,j})_{j\geq 0}$ is a sequence of i.i.d. random variables modeling successive thresholds. Otherwise $Q_{i,0} = 0$ and $(Q_{i,j})_{j\geq 1}$ is a sequence of i.i.d. random variables modeling successive thresholds. Then for $\alpha \in \{E, I_1, I_2, R\}$, RG_α denotes the sequences $(L_1^\alpha, \ldots, L_N^\alpha)$ of successive sojourn times in α, for each individual $i = 1, \ldots, N$. Finally, RG_M denotes the sequences $(M_1, \ldots, M_N)$ with $M_i = (M_{i,j})_{j\geq 1}$ a sequence of i.i.d. random variables distributed from the multinomial distribution over $\{E \to I_1, E \to I_2\}$ with parameter $\mathbf{p}$.

Algorithm 1 defines a deterministic function f such that, for all $t \geq 0$ and conditionally on $\{W^\theta(s), s < t)$, $W^\theta(t) = f(t, \theta, Z)$, with $\theta = (\beta, \mu_E, p, \mu_1, \mu_2, \delta)$ and $Z = (RG_Q, RG_E, RG_{I_1}, RG_{I_2}, RG_R, RG_M)$. Also, if θ is sampled from a random vector $\mathbf{X}$ with $\mathbf{X}$ and Z mutually independent, then (f, Z) defines a deterministic representation of $\left(\mathbf{X}, W^{\mathbf{X}}\right)$ in the sense that:

$$\left(\mathbf{X}, W^{\mathbf{X}}(\cdot)\right) \overset{\mathcal{D}}{=} (\mathbf{X}, f(\cdot, \mathbf{X}, Z)) .$$

As already mentioned, such a representation is a key stone for global sensitivity analysis (see Section 5 for more details).

Algorithm 1:

Data: $\xi_0 (= w(0)), T$
inputs : $\theta = (\beta, \mu_E, p, \mu_1, \mu_2, \delta)$,
$Z = \{RG_Q, RG_E, RG_{I_1}, RG_{I_2}, RG_R, RG_M\}$
output: $w = \{w(t); t \in [0, T]\}$

/* Initialization */
$t \leftarrow 0$, $w(t) \leftarrow \xi_0$, $P_S(t) \leftarrow 0$, $j \leftarrow 0$
$N \leftarrow w_S(t) + w_E(t) + w_{I_1}(t) + w_{I_2}(t) + w_R(t)$
/* Draw thresholds for initially susceptible individuals */
for *i such that* $X_i(0) = S$ **do**
  Draw $Q_{i,j}$ using the seed RG_Q
  $\widetilde{Q}_{i,j} \leftarrow Q_{i,j}$
end
/* Draw sojourn variables for individuals initially in E, I_1, I_2, R */
for $\alpha \in \{E, I_1, I_2, R\}$ **do**
  for *i such that* $X_i(0) = \alpha$ **do**
    Draw $L_{i,j}^\alpha$ using the seed RG_α
    $\tau_{i,j}^\alpha \leftarrow L_{i,j}^\alpha$
  end
end

while $t < T$ **do**

/* Pick the smallest threshold */

$\widetilde{Q}_{min} \leftarrow \min\{Q_{i,j} : i \text{ such that } X_i(t) = S\}$

$P_{deriv} \leftarrow (\beta/N) \times (w_{I_1}(t) + w_{I_2}(t))$

/* Compute putative next infection time */

$\tau^S \leftarrow t + \left(\widetilde{Q}_{min} - P_S(t)\right) / P_{deriv}$

/* Compute putative next exit time from other compartments */

for $\alpha \in \{E, I_1, I_2, R\}$ **do**

  $\tau^\alpha \leftarrow \min\{\tau^\alpha_{i,j} \mid i \text{ such that } X_i(t) = \alpha\}$

/* Compute next event time */

$\overline{\tau} \leftarrow \min\{\tau^\alpha : \alpha \in \{S, E, I_1, I_2, R\}\}$

/* Find next exit compartment */

$\alpha^* \leftarrow \text{argmin } \{\tau^\alpha \in \{S, E, I_1, I_2, R\}\}$

/* Compute pressure value at next exit time */

$P_S(\overline{\tau}) \leftarrow P_S(t) + P_{deriv} \times (\overline{\tau} - t)$

/* Identify the individual that makes the transition */

if $\alpha^* = S$ **then**

  $i^* \leftarrow \text{argmin } \{Q_{i,j} \mid X_i(t) = S\}$

  Delete $Q_{i^*,j}$

else if $\alpha^* \in \{E, I_1, I_2, R\}$ **then**

  $i^* \leftarrow \text{argmin } \{\tau^{\alpha^*}_{i,j} \mid X_i(t) = \alpha^*\}$

  Delete $\tau^{\alpha^*}_{i^*,j}$

/* Pick randomly destination compartment for a transition from E */

if $\alpha^* = E$ **then**

  Using the seed RG_M, draw (E, I_1) or (E, I_2) with probability p or $1 - p$ respectively.

else

  /* Destination compartment for transitions from other compartments */

  $\gamma^\star \leftarrow E$ if $\alpha^* = S$

  $\gamma^\star \leftarrow R$ if $\alpha^* = I_1$ or $\alpha^* = I_2$

  $\gamma^\star \leftarrow S$ if $\alpha^* = R$

/* Updates */

$w(\overline{\tau}) \leftarrow w(t) + \mathbf{u}_{\alpha^*,\gamma^\star}$

$t \leftarrow \overline{\tau}$, $j \leftarrow j + 1$

if $\gamma^\star = S$ **then**

  Draw $Q_{i^*,j}$ using the seed RG_Q

  $\widetilde{Q}_{i^*,j} \leftarrow P(t) + Q_{i^*,j}$

if $\gamma^\star \in \{E, I_1, I_2, R\}$ **then**

  Draw $L^{\gamma^\star}_{i^*,j}$ using the seed $RG_{\gamma^\star}$

  $\tau^{\gamma^\star}_{i^*,j} \leftarrow t + L^{\gamma^\star}_{i^*,j}$

In the framework of Proposition 1, the sequences of independent and identically distributed random variables Q_i, L_i^E, L_i^I and L_i^R are mutually independent and all sampled from an exponential distribution, and the resulting process $\{W(t)\ ,\ t \geq 0\}$ is a homogeneous continuous-time Markov chain. Now, if the sequences of independent and identically distributed random variables Q_i, L_i^E, L_i^I and L_i^R are mutually independent and if at least one of them is sampled from any non negative continuous probability distribution, different from the exponential distribution, and the other ones are sampled from an exponential distribution, then the corresponding process is non-Markovian.

4 Simulations

In this section, we present simulations of the SEI_1I_2RS model, obtained from Algorithm 1. We highlight differences between outputs depending on whether the model is Markovian or not. For this purpose, we consider the following epidemic scenario inspired from COVID-19 pandemic. The population is assumed to be composed of 2505 individuals with 5 initially exposed individuals (i.e. $W_E(0) = 5$). Population size is chosen arbitrary but not too large in order to avoid unnecessary computation burdens for our study which is methodological and do not claim to be a real case study. Regarding the epidemic characteristics, we assume that the incubation period $1/\mu_E$ fluctuates uniformly over $[4.5, 5.8]$ with mean 5.1 [Lauer et al., 2020]. Infectious periods $1/\mu_1$ and $1/\mu_2$ are respectively fixed to $1/\mu_1 = 5$ and $1/\mu_2 = 5$ according to Davies et al. [2020] (see Supplementary Table 1 in their paper). We then set the transmission rate to $\beta = 0.442$, so that for $1/\mu_1 = 5$ and $1/\mu_2 = 5$, the basic reproduction number $R_0 := \beta/\left(p \times \mu_1 + (1-p) \times \mu_2\right) = 2.21$, which seems reasonable. In addition, we set the proportion p to $18,1\%$ (see Prague et al. [2020]). In order to set the average period of loss of immunity $1/\delta$, we rely on Schuler et al. [2021]. In this paper, authors conclude that reinfection occurs from 90 to 180 days after recovery. Therefore, we set $1/\delta = 135$ corresponding to the mean of a uniform distribution over $[90, 180]$.

For these parameter values, two scenarios are considered for simulations, which correspond to the Markovian framework and a non-Markovian alternative. In both cases, sojourn times in a given compartment are independently and identically distributed (i.i.d.) from Gamma distributions. For a Gamma distribution, length of memory can be read in the shape parameter. Indeed, a Gamma distribution with shape parameter equal to one coincides with an exponential distribution; thus it is used to model the Markovian setting characterized by memoryless. To model scenarios with memory, we use in the following Gamma distributions with shape parameter greater than one. More precisely, we compare two scenarios. The Markovian scenario corresponds to exponential sampling of sojourn times in E, I_1, I_2 and R, with parameter chosen to fit respectively mean duration in E, I_1, I_2 and R. We also consider a non-Markovian scenario, with sojourn times in E, I_1 and I_2 sampled from a Gamma distribution with shape parameter equal to four (see [Davies

et al., 2020, Supplementary Table 1]) and scale parameter chosen to fit respectively mean duration in E, I_1 and I_2. The shape parameter for sojourn times in R is chosen equal to 1. We assume that the immune system of individuals lose completely immunity from one infection to the other. Tolerance thresholds are drawn in both Markovian and non-Markovian experiments from an exponential distribution with parameter 1.

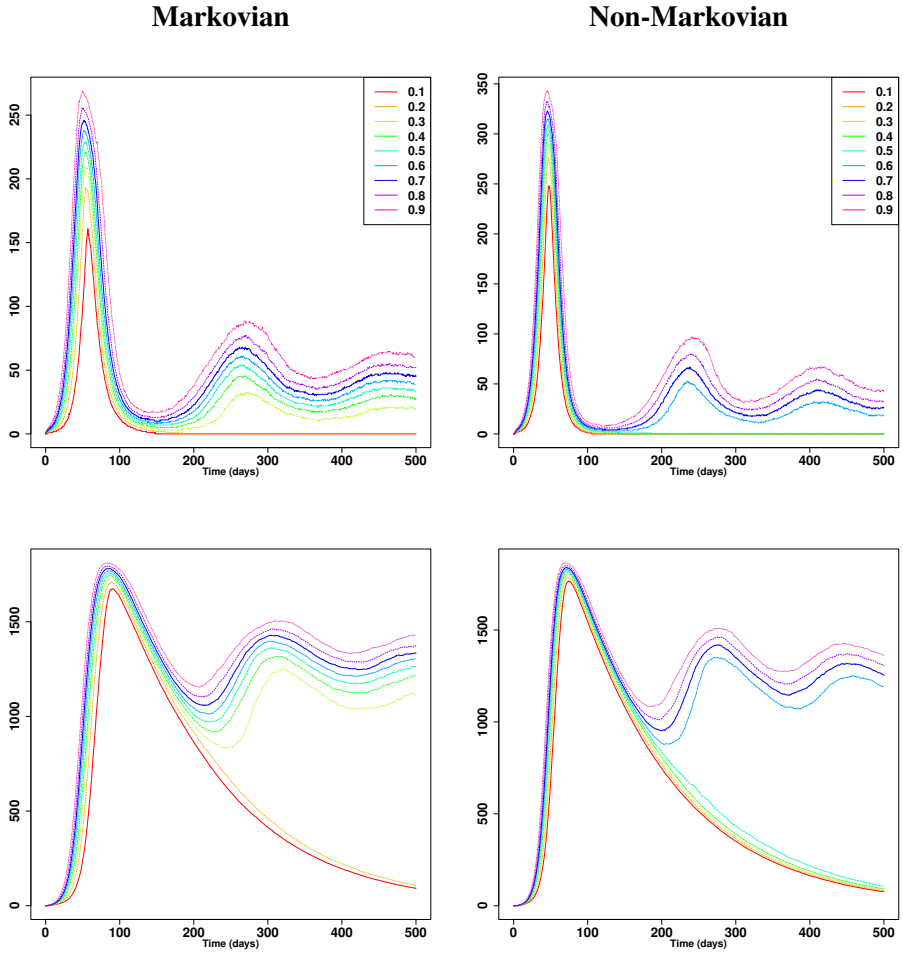

Figure 4: Evolution over time of quantiles of order 0.1 to 0.9 of the number of individuals in I_2 (top) and the one of recovered individuals (bottom). Quantiles are computed from 1000 independent trajectories in both Markovian (left) and non-Markovian (right) frameworks.

Using Algorithm 1, we simulate independently 50 sample paths or trajectories of the epidemic process over time period $[0, 500]$, for both Markovian and non-Markovian scenarios described above. Internal noise in Algorithm 1 is represented by $Z = (Q, L^E, L^{I_1}, L^{I_2}, L^R, M)$. Each realization $q^{(i)}$ (respectively $\ell^{E(i)}$,

${\ell^{I_1}}^{(i)}$, ${\ell^{I_2}}^{(i)}$, $\ell^{R(i)}$ and $m^{(i)}$) of Q (respectively L^E, L^{I_1}, L^{I_2}, L^R and M) is obtained from a pseudorandom number generator initialized with a random seed $z_1^{(i)}$ (respectively $z_2^{(i)}$, $z_3^{(i)}$, $z_4^{(i)}$, $z_5^{(i)}$ and $z_6^{(i)}$). Within each scenario, two trajectories differ only through the realization of Z. These simulations are provided in Figure C.1 in Appendix C. Differences can be observed from one scenario to the other. However, to limit the sampling impact of any conclusion, we simulate 1000 independent trajectories in both Markovian and non-Markovian frameworks, and plot in Figure 4 the evolution over time of quantiles of different order (ranged from 0.1 to 0.9) for the number of individuals in I_2 (top) and for the number of recovered individuals (bottom). On Figure 4, we read that out of our thousand simulations, at least 50% show a trend towards extinction for the non-Markovian scenario (top right), whereas this proportion falls to 20% for the Markovian scenario (top left). We also remark a significant difference in the peak for the dynamics of individuals in I_2 between the two scenarios (top line). The peak is significantly higher in the non-Markovian setting (top right) due to memory effect induced by the Gamma sampling of sojourn times.

With these simulations, we highlight the importance of modeling memory effects as they truly influence model outputs. To better account for the impact of memory effects, we perform in the next section a global sensitivity analysis.

5 Global sensitivity analysis of the SEI_1I_2RS model

This section is devoted to global sensitivity analysis of the SEI_1I_2RS model. We first review variance based sensitivity analysis in Section 5.1. Then in Section 5.2 we explain how to perform a global sensitivity analysis for a stochastic model based on a deterministic representation of the model, i.e. a deterministic function of the epidemic parameters and the variables modelling the internal noise. Finally in Section 5.3, we present numerical experiments for global sensitivity analysis of the SEI_1I_2RS model.

5.1 Sobol' indices

Let $\mathbf{X} = (X_1, \cdots, X_p)$ with $p \in \mathbb{N}^*$ be a random vector with known probability distribution. Let f be a real function taking p real arguments. Assume $\text{Var}(f(\mathbf{X})) < +\infty$. Moreover, suppose that $X_1, \cdots, X_p$ are mutually independent so that Sobol'-Hoeffding decomposition [Sobol', 1993, Hoeffding, 1948] yields:

$$\text{Var}\left(f(\mathbf{X})\right) = \sum_{u \subseteq \{1,\cdots,p\}, u \neq \emptyset} \text{Var}\left(f_u\left(\mathbf{X}_u\right)\right), \tag{12}$$

where, for each $u \subseteq \{1, \cdots, p\}$, f_u is a function of $X_u =: \{X_j, j \in u\}$ such that $\mathbb{E}\left[f_u\left(\mathbf{X}_u\right) \mid \mathbf{X}_v\right] = 0$ for every $v \subsetneq u$. Based on (12), Sobol' [1993] introduced variance-based sensitivity indices, nowadays known as Sobol' indices. In this paper we focus on first-order and total Sobol' indices.

First-order Sobol' indices Let u be a nonempty subset of $\{1, \cdots p\}$. The first-order Sobol' index of $\mathbf{X}_u$ is defined as:

$$S_{\mathbf{X}_u} = \frac{\text{Var}\left(\mathbb{E}\left[f(\mathbf{X}) \mid \mathbf{X}_u\right]\right)}{\text{Var}(f(\mathbf{X}))}. \tag{13}$$

Such an index is a ratio of the variance of the part of the output due only to the variation of $\mathbf{X}_u$ and the global variance. It is the so-called main effect of $\mathbf{X}_u$.

Total Sobol' indices The total Sobol index of $\mathbf{X}_u$ is defined as

$$ST_{\mathbf{X}_u} = 1 - \frac{\text{Var}\left(\mathbb{E}\left[f(\mathbf{X}) \mid \mathbf{X}_{\sim u}\right]\right)}{\text{Var}(f(\mathbf{X}))}, \tag{14}$$

where $\mathbf{X}_{\sim u} = \{X_j, j \notin u\}$. This index summarizes the main effect of $\mathbf{X}_u$ as well as its interactions with all other inputs of the model. Indeed, for any $j \in \{1, \ldots, p\}$, we have:

$$ST_j = \sum_{u \subseteq \{1,\ldots,p\} \text{ such that } j \in u} S_{\mathbf{X}_u}.$$

Remark 2. *The definition of first-order and total Sobol' indices can be extended to models with vectorial or functional output. For a vectorial output $(Y_1, \ldots, Y_p)$, and for any input parameter X_i, it is possible to compute the so-called aggregated first-order (or total) Sobol' index by computing the weighted sum of first-order (or total) Sobol' indices for each scalar output Y_j weighted by the variance of Y_j. We refer to Lamboni et al. [2011], Gamboa et al. [2014] for the explicit definition. If the output of the model of interest is a function of time, it can be reduced to a vectorial output through discretization of time. As an alternative, first-order and total indices defined in Equations* (13) *and* (14) *can be computed at each discretized time.*

5.2 Deterministic representation of stochastic models

Let $\mathcal{X}, \mathcal{Y}$ and $\mathcal{Z}$ be three nonempty measurable spaces. A stochastic model g is defined as follows. For each set of input parameters $x \in \mathcal{X}$, the model output $g(x, \cdot)$ is a random variable with values on $\mathcal{Y}$, so that a realization of such an output is under the form $g(x, \omega)$, where ω belongs to some σ-algebra. In the paradigm of global sensitivity analysis, uncertain parameters are modeled by a random vector $\mathbf{X}$. In the following, we make the usual assumption that for every $x \in \mathcal{X}$, $\mathbf{X}$ and $g(x, \cdot)$ are mutually independent. In other words, uncertain parameters are independent from internal noise.

Definition 1. *[Kouye et al., 2024, Section 2.2] Let $f : \mathcal{X} \times \mathcal{Z} \to \mathcal{Y}$ be a deterministic function. Let Z be a random variable with values in $\mathcal{Z}$, whose probability distribution is known, with Z and $\mathbf{X}$ mutually independent. Then (f, Z) is said to be a deterministic representation of the stochastic model g if $(\mathbf{X}, g(\mathbf{X}, \cdot)) \overset{\mathcal{D}}{=} (\mathbf{X}, f(\mathbf{X}, Z))$, where $\overset{\mathcal{D}}{=}$ means equality in distribution. The random variable Z controls the internal noise of the stochastic model.*

Remark 3. *In particular, if for every set of input parameters* $x \in \mathcal{X}$, $g(x,\cdot) \overset{\mathcal{D}}{=} f(x,Z)$, *then* $(\mathbf{X}, g(\mathbf{X},\cdot)) \overset{\mathcal{D}}{=} (\mathbf{X}, f(\mathbf{X},Z))$.

A strategy to perform global sensitivity analysis of a stochastic model g is to first identify a deterministic representation (f,Z) of g and then to compute Sobol' indices as in Section 5.1 by considering the deterministic model f with augmented input vector $(\mathbf{X},Z)$. Let $j \in \{1,\dots,p\}$. Then the impact of the different sources of uncertainty on model output are measured by the computation of Sobol' indices: S_{X_j} quantifies the impact of X_j, S_Z measures the impact of Z, the variable that controls the stochasticity of the model. We also calculate the total Sobol' index ST_{X_j}, in order to quantify the impact of X_j alone or in interaction with other epidemic parameters $X_{j'}$, $j' \neq j$ or with Z, and the total Sobol' index ST_Z.

In the next section, we apply this methodology to perform a global sensitivity analysis of simulations of the SEI_1I_2RS model obtained from Algorithm 1 presented in Section 3.3.

5.3 Sensitivity analysis

In this section, we perform a global sensitivity analysis for two different quantities of interest in the SEI_1I_2RS model. We compare the results obtained in a Markovian framework and in a non-Markovian framework. As the results are different, this highlights the importance of taking into account the memory (if any) effect when modeling an epidemic. Throughout this section, we can think about I_1 and I_2 as the compartments for asymptomatic and symptomatic infectious individuals. Each epidemic parameter is sampled from a uniform distribution whose support is set according to the literature on COVID-19. Indeed, even if we do not pretend to consider a real case study, we aim to consider a realistic setting.

Two quantities of interest are under study. The first quantity is the dynamics over time of infected individuals, namely $W_E(t),\ t \in [0,500]$. The second quantity is the moment the peak of the number individuals in I_2 is reached, namely $\text{argmax}_{t\in[0,500]} W_{I_2}(t)$. In the context of COVID-19, this corresponds to the moment the peak of the number of symptomatic individuals is reached. In practice, this quantity is of particular interest. It is linked to the overload of the medical system, which proved to be a major problem during the COVID-19 epidemic.

Numerical setting

In our numerical experiments, we consider the Markovian and the non-Markovian frameworks described in Section 4. The size of the population, as well as initial conditions, are also chosen as in Section 4. Epidemic parameters β, p, μ_E, μ_1, μ_2 and δ are sampled uniformly (the range of variation for each parameter is described in Table 4).

Parameter Name	Nominal value	Range of variation
β	0.442	$[0.23, 0.53]$
$1/\mu_E$	5.1 days	$[4.5, 5.8]$
p	0.181	$[0.167, 0.192]$
$1/\mu_1$	5 days	$[5, 8]$
$1/\mu_2$	5 days	$[5, 8]$
$1/\delta$	135 days	$[90, 180]$

Table 4: Nominal values and range of variation of uncertain epidemic parameters

To sample the variable Z, that controls internal noise, we sample the seed of pseudo-random number generator uniformly in $\{1, \cdots, 10^9\}$. We compute Sobol' indices by using the function `soboljansen` of the R-package `sensitivity` [Iooss et al., 2020]. Each evaluation of Sobol' indices uses a n-sample of $(p, \mu_E, \mu_1, \mu_2, \delta, Z)$ with $n = 2500$. The time-dependent output $W_E(t)$, $t \in [0; 500]$ is evaluated on a regular grid $t_0 = 0, \ldots, t_m = 500$ with $m = 1000$. Then we can compute Sobol' indices at each time t_i of the grid, or alternatively compute aggregated Sobol' indices (see Remark 2 for a definition). To account for variability due to input sampling, each Sobol' index is evaluated $N_{rep} = 30$ times.

Results for $\mathbf{W_E(t)}$, $\mathbf{t \in [0, 500]}$ Figure 5 shows the evolution over time of first-order (top) and total (bottom) Sobol' indices. For both Markovian and non-Markovian scenarios we consider, we note that, except during the first phase of the epidemic, when the impact of internal noise Z is the strongest, β is the most important input. Only slight differences can be observed on the evolution of the influence of β from one scenario to the other. Though, a significant difference can be observed for the impact of internal noise Z between both scenarios. We observe that Z has much more impact in the Markovian scenario. This is even clearer on Figure 6, where we have plotted first-order and total aggregated Sobol' indices to summarize the influence of inputs on the dynamical output $W_E(t)$, $t \in [0, 500]$.

Figure 5: Time evolution of the median of a sample of $N_{rep} = 30$ realizations of first-order (top) and total (bottom) Sobol' index estimators in Markovian (left) and non-Markovian (right) frameworks. The estimators are built from an input sample of size $n = 2500$.

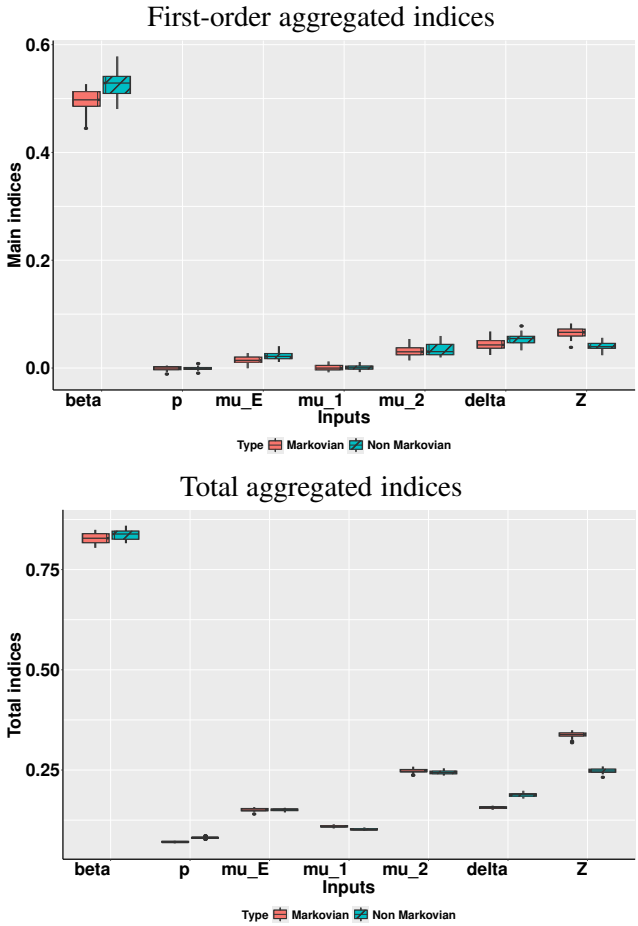

Figure 6: Boxplots obtained from $N_{rep} = 30$ realizations of first-order (left) and total (right) aggregated Sobol' index estimators in our Markovian (red) and non-Markovian (blue) scenarios. The estimators are built from an input sample of size $n = 2500$.

Results for $\mathbf{argmax}_{\mathbf{t}\in[\mathbf{0,500}]}\mathbf{W}_{\mathbf{I_2}}(\mathbf{t})$ On Figure 7 are plotted the results for first-order and total Sobol' index estimation.

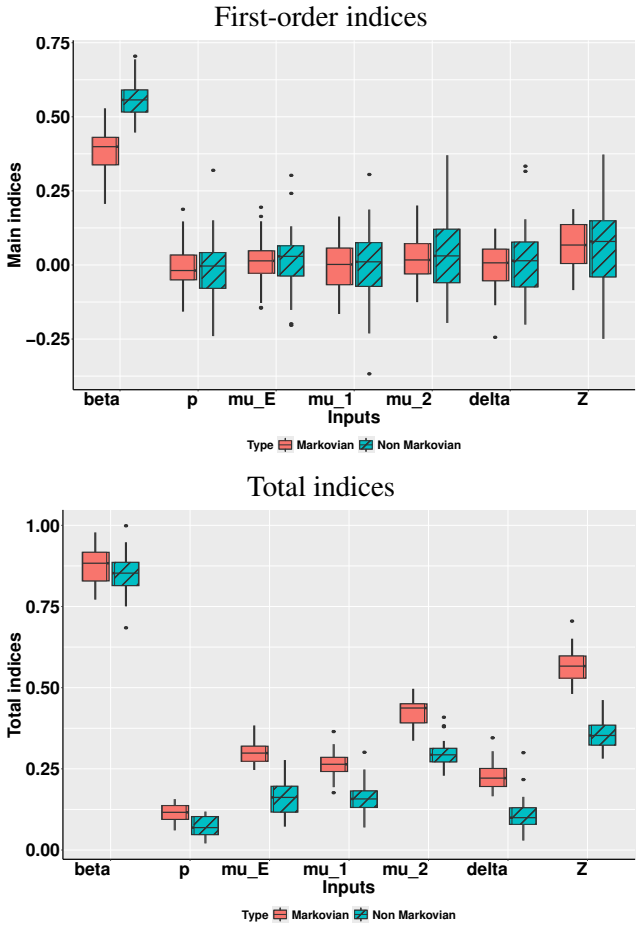

Figure 7: Boxplots obtained from $N_{rep} = 30$ realizations of first-order (top) and total (bottom) Sobol' index estimators in our Markovian (red) and non-Markovian (blue) scenarios. The estimators are built from an input sample of size $n = 2500$.

We note that β is always the most important input for this quantity of interest. From the plots of first-order indices, β explains a little bit more than 37.5% of the variance of the output in the Markovian scenario, while in the non-Markovian one, this proportion increases up to 55%. Considering the total Sobol' indices, the dominance of the influence of β is confirmed. In addition, there are significant differences in the total effects of all other input parameters between the Markovian scenario and the non-Markovian one. The larger difference is observed in the total effect of Z. The impact of internal noise is greater in the Markovian scenario than in the non-Markovian one. To highlight the pure interaction effects, we compute for each input parameter the value of the total Sobol' index minus the one of the

first-order Sobol' index. Corresponding boxplots are plotted in Figure 8.

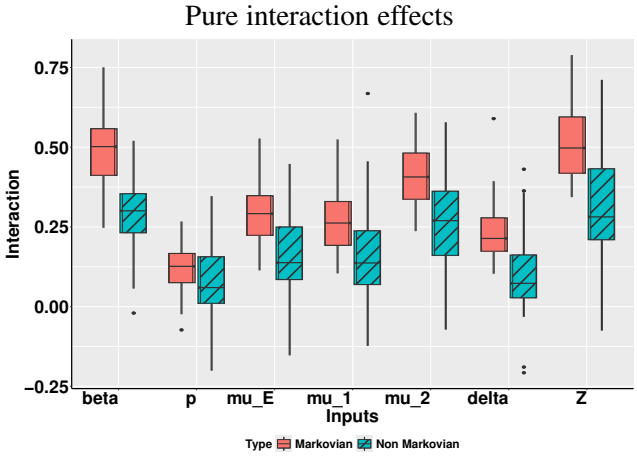

Figure 8: Boxplots obtained from $N_{rep} = 30$ realizations of pure interaction effects in our Markovian (red) and non-Markovian (blue) scenarios. The estimators are built from an input sample of size $n = 2500$.

6 Conclusion

In epidemiology, compartmental modeling plays a key role in the quantitative study and understanding of the spread of epidemics. Two frameworks are available to modelers in stochastic modeling: the Markovian framework and the non-Markovian framework. The Markovian framework consists in using Markovian processes such as continuous-time Markov chains, i.e. memoryless processes, to model epidemics. The advantage of this framework is that an arsenal of mathematical tools is available to facilitate theoretical analysis. On the other hand, the non-Markovian framework is more difficult to deal with theoretically, but it is more general and more realistic, as it does not rely on the absence of memory. The choice of one of these frameworks has consequences for the lessons or conclusions that can be drawn from the study of the resulting epidemic process.

In this paper, we use sensitivity analysis to highlight the differences between different scenarios, with or without memory effects. These differences appear in the influence of epidemic parameters but also in the influence of intrinsic noise on epidemic dynamics. The key stone for our sensitivity analysis is the extension of Sellke construction to more complex models. Note that, although we focused the presentation on the SEI_1I_2RS model, our construction easily extends to any closed-population compartmental model with loops and parallel compartments. Our extension of Sellke construction inherits its ability to simulate exact trajectories of a same model in a Markovian context (by sampling exponential dis-

tributions for sojour times) but also in a non-Markovian context (by replacing exponential sampling by, e.g., Gamma sampling).

In future work, it would be interesting to compare simulation results obtained from Algorithm 1 based on our extension of Sellke construction to other algorithms dedicated to non-Markovian dynamics (e.g. Vestergaard and Génois [2015], Boguñá et al. [2014], Masuda and Rocha [2018]). In particular, robustness to internal noise may vary from one algorithm to the other, with an impact on global sensitivity analysis results, as studied in Kouye et al. [2024] in the Markovian context.

References

Håkan Andersson and Tom Britton. *Stochastic Epidemic Models and Their Statistical Analysis*, volume 151 of *Lecture Notes in Statistics*. Springer New York, New York, NY, 2000. ISBN 978-0-387-95050-1 978-1-4612-1158-7. doi: 10.1007/978-1-4612-1158-7.

Marian Boguñá, Luis F. Lafuerza, Raúl Toral, and M. Ángeles Serrano. Simulating non-Markovian stochastic processes. *Phys. Rev. E*, 90:042108, Oct 2014. doi: 10.1103/PhysRevE.90.042108.

Tobias Brett and Tobias Galla. Stochastic Processes with Distributed Delays: Chemical Langevin Equation and Linear-Noise Approximation. *Phys. Rev. Lett.*, 110:250601, Jun 2013. doi: 10.1103/PhysRevLett.110.250601.

Tom Britton and Etienne Pardoux. *Chapter 3 General Closed Models*, pages 43–57. Springer International Publishing, Cham, 2019a. ISBN 978-3-030-30900-8. doi: 10.1007/978-3-030-30900-8_3.

Tom Britton and Etienne Pardoux, editors. *Stochastic Epidemic Models with Inference*, volume 2255 of *Lecture Notes in Mathematics*. Springer International Publishing, Cham, 2019b. ISBN 978-3-030-30899-5 978-3-030-30900-8. doi: 10.1007/978-3-030-30900-8.

Nicholas G. Davies, Petra Klepac, Yang Liu, Kiesha Prem, Mark Jit, CMMID COVID-19 working group, Carl A. B. Pearson, Billy J. Quilty, Adam J. Kucharski, Hamish Gibbs, Samuel Clifford, Amy Gimma, Kevin Van Zandvoort, James D. Munday, Charlie Diamond, W. John Edmunds, Rein M. G. J. Houben, Joel Hellewell, Timothy W. Russell, Sam Abbott, Sebastian Funk, Nikos I. Bosse, Yueqian Fiona Sun, Stefan Flasche, Alicia Rosello, Christopher I. Jarvis, and Rosalind M. Eggo. Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nature Medicine*, 26(8):1205–1211, August 2020. ISSN 1078-8956, 1546-170X. doi: 10.1038/s41591-020-0962-9.

Francesco Di Lauro, Wasiur R. KhudaBukhsh, István Z. Kiss, Eben Kenah, Max Jensen, and Grzegorz A. Rempała. Dynamic survival analysis for non-Markovian epidemic models. *Journal of The Royal Society Interface*, 19(191): 20220124, 2022. doi: 10.1098/rsif.2022.0124.

Stewart N. Ethier and Thomas G. Kurtz. *Markov processes – characterization and convergence*, chapter 4, 6. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986. ISBN 0-471-08186-8.

Fabrice Gamboa, Alexandre Janon, Thierry Klein, and Agnès Lagnoux. Sensitivity analysis for multidimensional and functional outputs. *Electronic Journal of Statistics*, 8(1):575 – 603, 2014. doi: 10.1214/14-EJS895.

Daniel T Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403–434, 1976. ISSN 0021-9991. doi: https://doi.org/10.1016/0021-9991(76)90041-3.

Gerrit Großmann, Michael Backenköhler, and Verena Wolf. Why ODE models for COVID-19 fail: Heterogeneity shapes epidemic dynamics. *medRxiv*, 2021. doi: 10.1101/2021.03.25.21254292.

Wassily Hoeffding. A Class of Statistics with Asymptotically Normal Distribution. *The Annals of Mathematical Statistics*, 19(3):293 – 325, 1948. doi: 10.1214/aoms/1177730196.

Thomas House. Non-markovian stochastic epidemics in extremely heterogeneous populations. *Math. Model. Nat. Phenom.*, 9(2):153–160, 2014. doi: 10.1051/mmnp/20149210.

Bertrand Iooss, Sebastien Da Veiga, Alexandre Janon, Gilles Pujol, with contributions from Baptiste Broto, Khalid Boumhaout, Thibault Delage, Reda El Amri, Jana Fruth, Laurent Gilquin, Joseph Guillaume, Loic Le Gratiet, Paul Lemaitre, Amandine Marrel, Anouar Meynaoui, Barry L. Nelson, Filippo Monari, Roelof Oomen, Oldrich Rakovec, Bernardo Ramos, Olivier Roustant, Eunhye Song, Jeremy Staum, Roman Sueur, Taieb Touati, and Frank Weber. *sensitivity: Global Sensitivity Analysis of Model Outputs*, 2020. R package version 1.19.0.

William Ogilvy Kermack and Anderson G McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721, 1927.

Henri Mermoz Kouye, Gildas Mazo, Clémentine Prieur, and Elisabeta Vergu. Performing global sensitivity analysis on simulations of a continuous-time Markov chain model motivated by epidemiology. *Computational and Applied Mathematics*, 43(7):409, 2024. ISSN 1807-0302. doi: 10.1007/s40314-024-02897-y.

Matieyendou Lamboni, Monod Hervé, and Makowski David. Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models. *Reliability Engineering and System Safety*, 96(4):450 – 459, 2011. ISSN 0951-8320. doi: https://doi.org/10.1016/j.ress.2010.12.002.

Stephen A. Lauer, Kyra H. Grantz, Qifang Bi, Forrest K. Jones, Qulu Zheng, Hannah R. Meredith, Andrew S. Azman, Nicholas G. Reich, and Justin Lessler. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Annals of Internal Medicine*, 172(9):577–582, May 2020. ISSN 0003-4819, 1539-3s704. doi: 10.7326/M20-0504.

O. P. Le Maître, O. M. Knio, and A. Moraes. Variance decomposition in stochastic simulators. *The Journal of Chemical Physics*, 142(24):244115, 2015. doi: 10.1063/1.4922922.

Naoki Masuda and Luis E. C. Rocha. A Gillespie Algorithm for Non-Markovian Stochastic Processes. *SIAM Review*, 60(1):95–115, 2018. doi: 10.1137/16M1055876.

Cameron Nowzari, Masaki Ogura, Victor M. Preciado, and George J. Pappas. A general class of spreading processes with non-Markovian dynamics. In *2015 54th IEEE Conference on Decision and Control (CDC)*, pages 5073–5078, 2015. doi: 10.1109/CDC.2015.7403013.

Matthew J. Penn, Daniel J. Laydon, Joseph Penn, Charles Whittaker, Christian Morgenstern, Oliver Ratmann, Swapnil Mishra, Mikko S. Pakkanen, Christl A. Donnelly, and Samir Bhatt. Intrinsic randomness in epidemic modelling beyond statistical uncertainty. *Communications Physics*, 6(1):146, June 2023. ISSN 2399-3650. doi: 10.1038/s42005-023-01265-2.

Mélanie Prague, Linda Wittkop, Quentin Clairon, Dan Dutartre, Rodolphe Thiébaut, and Boris P. Hejblum. Population modeling of early covid-19 epidemic dynamics in french regions and estimation of the lockdown impact on infection rate. *medRxiv*, 2020. doi: 10.1101/2020.04.21.20073536.

Gesine Reinert. The Asymptotic Evolution of the General Stochastic Epidemic. *The Annals of Applied Probability*, 5(4):1061 – 1086, 1995. doi: 10.1214/aoap/1177004606.

M. Saeedian, M. Khalighi, N. Azimi-Tafreshi, G. R. Jafari, and M. Ausloos. Memory effects on epidemic evolution: The susceptible-infected-recovered epidemic model. *Phys. Rev. E*, 95:022409, Feb 2017. doi: 10.1103/PhysRevE.95.022409.

Charles F Schuler, Carmen Gherasim, Kelly O'Shea, David M Manthei, Jesse Chen, Cristyn Zettel, Jonathan P Troost, Andrew A Kennedy, Andrew W Tai, Donald A Giacherio, Riccardo Valdez, James L Baldwin, and James R Baker.

Mild sars-cov-2 illness is not associated with reinfections and provides persistent spike, nucleocapsid, and virus-neutralizing antibodies. *Microbiology spectrum*, 9(2):e0008721, October 2021. ISSN 2165-0497. doi: 10.1128/spectrum.00087-21.

Thomas Sellke. On the asymptotic distribution of the size of a stochastic epidemic. *Journal of Applied Probability*, 20(2):390–394, 1983. doi: 10.2307/3213811.

I. M. Sobol'. Sensitivity analysis for non-linear mathematical models. *Mathematical Modelling and Computational Experiment*, 1:407–414, 1993.

Mircea T. Sofonea, Bastien Reyné, Baptiste Elie, Ramsès Djidjou-Demasse, Christian Selinger, Yannis Michalakis, and Samuel Alizon. Memory is key in capturing COVID-19 epidemiological dynamics. *Epidemics*, 35:100459, 2021. ISSN 1755-4365. doi: https://doi.org/10.1016/j.epidem.2021.100459.

G. Streftaris and G. J. Gibson. Non-exponential tolerance to infection in epidemic systems–modeling, inference, and assessment. *Biostatistics*, 13(4):580–593, September 2012. ISSN 1465-4644, 1468-4357. doi: 10.1093/biostatistics/kxs011.

P. Van Mieghem and R. van de Bovenkamp. Non-Markovian Infection Spread Dramatically Alters the Susceptible-Infected-Susceptible Epidemic Threshold in Networks. *Phys. Rev. Lett.*, 110:108701, Mar 2013. doi: 10.1103/PhysRevLett.110.108701.

Christian L. Vestergaard and Mathieu Génois. Temporal Gillespie Algorithm: Fast Simulation of Contagion Processes on Time-Varying Networks. *PLOS Computational Biology*, 11(10):1–28, 10 2015. doi: 10.1371/journal.pcbi.1004579.

A Proof of Proposition 1

Let $t \geq 0$ and denote by w the current state of W with $\mathbb{P}(W(t) = w) > 0$. Let $\Delta > 0$. In the following, Δ will tend to zero.

Infection rate: The probability of infection between t and $t+\Delta$ is:

$$\begin{aligned}
&\mathbb{P}\left(W(t+\Delta)-W(t)=\mathbf{u}_{(S,E)} \mid W(t)=w\right)\\
&=\textstyle\sum_{i:X_i(t)=S}\mathbb{P}\left(Q_{i,j+1}+P(\eta^S_{i,j})\le P(t+\Delta) \mid Q_{i,j+1}+P(\eta^S_{i,j})>P(t)\right)\\
&=\textstyle\sum_{i:X_i(t)=S}\Big(1-\mathbb{P}\Big(Q_{i,j+1}+P(\eta^S_{i,j})>P(t)+\frac{\beta}{N}\left(W_{I_1}(t)+W_{I_2}(t)\right)\Delta+o(\Delta)\ \mid\\
&\qquad\qquad Q_{i,j+1}+P(\eta^S_{i,j})>P(t)\Big)\Big)\\
&=\textstyle\sum_{i:X_i(t)=S}\left(1-\exp\left(-\frac{\beta}{N}\left(W_{I_1}(t)+W_{I_2}(t)\right)\Delta+o(\Delta)\right)\right)\\
&=W_S(t)\left(1-\exp\left(-\frac{\beta}{N}\left(W_{I_1}(t)+W_{I_2}(t)\right)\Delta+o(\Delta)\right)\right).
\end{aligned}$$

As $\Delta\to 0$, $W_S(t)\left(1-\exp\left(-\frac{\beta}{N}\left(W_{I_1}(t)+W_{I_2}(t)\right)\Delta+o(\Delta)\right)\right)/\Delta\to\frac{\beta}{N}W_S(t)(W_{I_1}(t)+W_{I_2}(t))=\frac{\beta}{N}W_S(t)W_I(t)$.

Transitions $\mathbf{E}\to\mathbf{I_1}$: The probability $\mathbb{P}\left(W(t+\Delta)-W(t)=\mathbf{u}_{(E,I_1)} \mid W(t)=w\right)$ computes as

$$\begin{aligned}
&\mathbb{P}\left(W(t+\Delta)-W(t)=\mathbf{u}_{(E,I_1)} \mid W(t)=w\right)\\
&=\textstyle\sum_{i:X_i(t)=E}\mathbb{P}\left(L^E_{i,j+1}+\eta^E_{i,j}\le t+\Delta, M_{i,j+1}=(E,I_1) \mid L^E_{i,j+1}+\eta^E_{i,j}>t\right)\\
&=\textstyle\sum_{i:X_i(t)=E}\mathbb{P}\left(M_{i,j+1}=(E,I_1)\right)\left(1-\mathbb{P}\left(L^E_{i,j+1}+\eta^E_{i,j}>t+\Delta \mid L^E_{i,j+1}+\eta^E_{i,j}>t\right)\right)\\
&=\textstyle\sum_{i:X_i(t)=E}p\left(1-\exp(-\mu_E\Delta)\right)=pW_E(t)\left(1-\exp(-\mu_E\Delta)\right).
\end{aligned}$$

As $\Delta\to 0$, $pW_E(t)\left(1-\exp(-\mu_E\Delta)\right)/\Delta\to p\mu_E W_E(t)$. Computations for transitions $\mathbf{E}\to\mathbf{I_2}$ are similar.

Other transitions $(\boldsymbol{\alpha_1},\boldsymbol{\alpha_2})$ We have:

$$\begin{aligned}
&\mathbb{P}\left(W(t+\Delta)-W(t)=\mathbf{u}_{(\boldsymbol{\alpha_1},\boldsymbol{\alpha_2})} \mid W(t)=w\right)\\
&=\textstyle\sum_{i:X_i(t)=\alpha_1}\mathbb{P}\left(L^{\alpha_1}_{i,j+1}+\eta^{\alpha_1}_{i,j}\le t+\Delta \mid L^{\alpha_1}_{i,j+1}+\eta^{\alpha_1}_{i,j}>t\right)\\
&=\textstyle\sum_{i:X_i(t)=\alpha_1}\left(1-\mathbb{P}\left(L^{\alpha_1}_{i,j+1}+\eta^{\alpha_1}_{i,j}>t+\Delta \mid L^{\alpha_1}_{i,j+1}+\eta^{\alpha_1}_{i,j}>t\right)\right)\\
&=\textstyle\sum_{i:X_i(t)=\alpha_1}\left(1-\exp(-\mu_{\alpha_1}\Delta)\right)=W_{\mu_1}(t)\left(1-\exp(-\mu_{\alpha_1}\Delta)\right),
\end{aligned}$$

with $\mu_{\alpha_1} = \begin{cases} \mu_1 & \text{if } \alpha_1 = I_1, \\ \mu_2 & \text{if } \alpha_1 = I_2, \\ \delta & \text{if } \alpha_1 = R. \end{cases}$

B Proof of (10) in Section 3.3

Let $\alpha \in \{S, E, I, R\}$.

For $j \geq 0$, $X_i(0) > \alpha$,

$$\Delta^{\alpha}_{i,j+1} = \sum_{\alpha < \gamma \leq X_i(0)} L^{\gamma}_{i,j} + \sum_{\gamma > X_i(0)} L^{\gamma}_{i,j+1} + \sum_{\gamma \leq \alpha} L^{\gamma}_{i,j+1}\,. \tag{15}$$

For $j \geq 0$, $X_i(0) = \alpha$,

$$\Delta^{\alpha}_{i,j+1} = \sum_{\gamma} L^{\gamma}_{i,j+1}\,. \tag{16}$$

For $j \geq 0$, $X_i(0) < \alpha$,

$$\Delta^{\alpha}_{i,j+1} = \sum_{\gamma > \alpha} L^{\gamma}_{i,j} + \sum_{\gamma \leq X_i(0)} L^{\gamma}_{i,j} + \sum_{X_i(0) < \gamma \leq \alpha} L^{\gamma}_{i,j+1}\,. \tag{17}$$

We now apply the above equations to $\alpha = S$.

For $j \geq 0$, $X_i(0) > S$,

$$\tau^{S}_{i,j+1} - \tau^{S}_{i,j} = \sum_{S < \gamma \leq X_i(0)} L^{\gamma}_{i,j} + \sum_{\gamma > S} L^{\gamma}_{i,j+1} + L^{S}_{i,j+1}\,. \tag{18}$$

For $j \geq 0$, $X_i(0) = S$,

$$\tau^{S}_{i,j+1} - \tau^{S}_{i,j} = \sum_{\gamma \neq S} L^{\gamma}_{i,j+1} + L^{S}_{i,j+1}\,. \tag{19}$$

Now, to prove (10),

- $X_i(0) = S$, $\alpha \neq S$, combine (17) and (19);
- $X_i(0) > S$ and $\alpha < X_i(0)$, combine (15) and (18);
- $X_i(0) > S$ and $\alpha = X_i(0)$, combine (16) and (18);
- $X_i(0) > S$ and $\alpha > X_i(0)$, combine (17) and (18).

C Plots

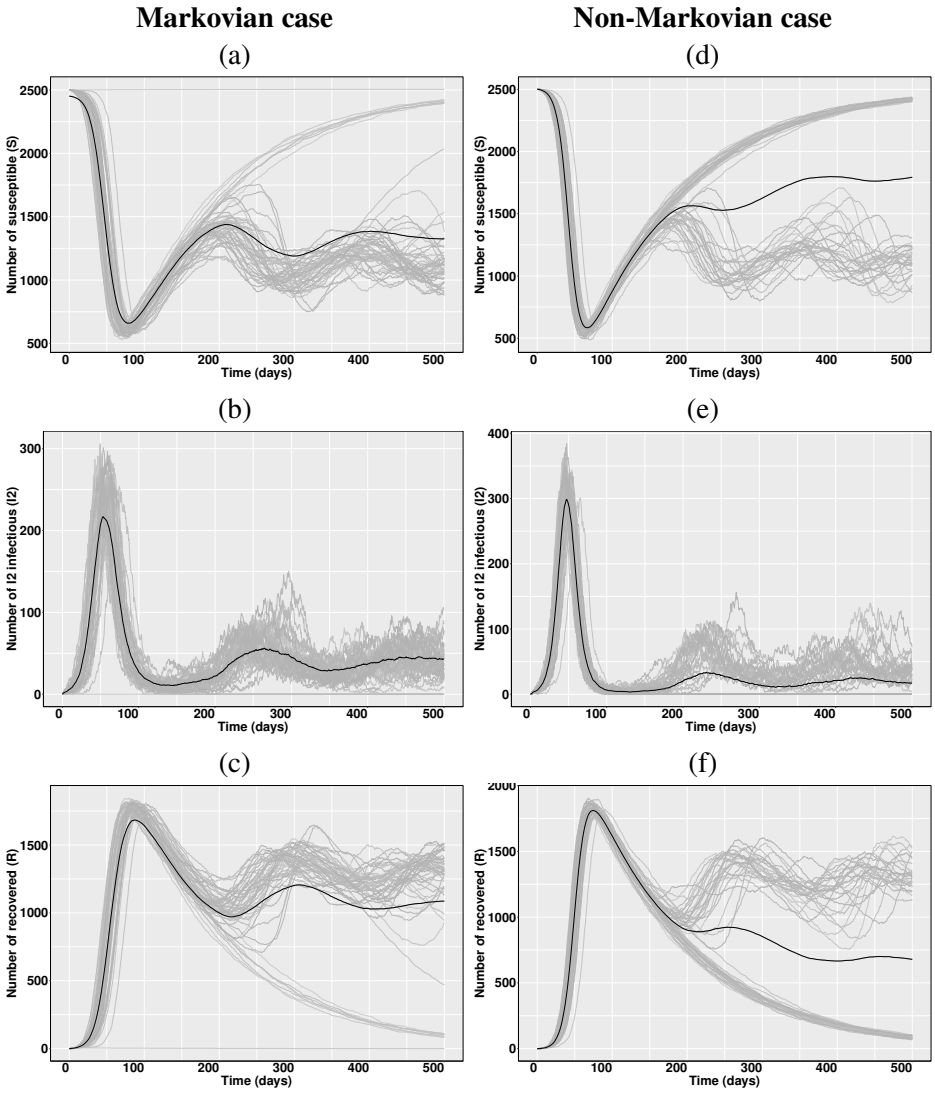

Figure C.1: 50 independent trajectories over time period $[0, 500]$ of number of individuals in compartments S, I_1, I_2 and R in Markovian framework (a, b, c) and non-Markovian framework (d, e, f).