



**HAL**  
open science

# Exploiting deterministic algorithms to perform global sensitivity analysis for continuous-time Markov chain compartmental models with application to epidemiology

Henri Mermoz Kouye, Gildas Mazo, Clémentine Prieur, Elisabeta Vergu

## ► To cite this version:

Henri Mermoz Kouye, Gildas Mazo, Clémentine Prieur, Elisabeta Vergu. Exploiting deterministic algorithms to perform global sensitivity analysis for continuous-time Markov chain compartmental models with application to epidemiology. 2022. hal-03565729v1

**HAL Id: hal-03565729**

**<https://hal.science/hal-03565729v1>**

Preprint submitted on 14 Feb 2022 (v1), last revised 22 Jul 2024 (v6)

**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.

# Exploiting deterministic algorithms to perform global sensitivity analysis for continuous-time Markov chain compartmental models with application to epidemiology

Henri Mermoz KOUYE<sup>(1)</sup>, Gildas MAZO<sup>(1)</sup>, Clémentine PRIEUR<sup>(2)</sup>  
and Elisabeta VERGU<sup>(1)</sup>

<sup>(1)</sup>Univ. Paris-Saclay, INRAE, MaIAGE, 78350, Jouy-en-Josas, France

<sup>(2)</sup>Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, LJK, 38000, Grenoble, France

February 14, 2022

## Abstract

In this paper, we develop an approach of global sensitivity analysis for compartmental models based on continuous-time Markov chains. We propose to measure the sensitivity of quantities of interest by representing the Markov chain as a deterministic function of the uncertain parameters and a random variable with known distribution modeling intrinsic randomness. This representation is exact and does not rely on meta-modeling. An application to a SARS-CoV-2 epidemic model is included to illustrate the practical impact of our approach.

**Keywords:** Compartmental models, continuous-time Markov chains, Gillespie algorithm, stochastic models, deterministic, global sensitivity analysis.

## 1 Introduction

The increasing use of mathematical modeling leads to an enhanced complexity of computational models. These models can be seen as mappings which take inputs and return random or deterministic outputs. If the output is random, two evaluations of the model at the same input generate different realizations: the model is said to be stochastic. Otherwise, the model is deterministic.

Stochastic models are often used in epidemiology. Indeed, in order to study and control the spread of infectious diseases in populations (humans, animals or plants), stochastic compartmental models enable to describe epidemic dynamics by incorporating randomness associated with biological and contact events. They consist in dividing the population into disjoint groups with respect to the different health statuses that are considered. The groups form the compartments of the model. As health statuses of individuals change over time, there are transitions between compartments. These transitions occur at random times and depend on numerous uncertain parameters (also called the input hereafter). Even by fixing uncertain parameters, the number of individuals in each compartment varies randomly over time. Therefore, the corresponding process is

stochastic and under some modeling assumptions it is a continuous-time Markov chain (CTMC).

Very often, transition parameters of CTMC are poorly known. In order to better characterize and predict epidemic spread and assess corresponding control strategies, it is important to identify key parameters of the infection spread accounting parameter uncertainty. For this purpose, global sensitivity analysis (GSA) can be used. GSA enables to assess influence of uncertain parameters on the model output. However, performing GSA for stochastic model output is challenging. Unlike deterministic models, stochastic models include two sources of uncertainty: parameter uncertainty and intrinsic randomness. Indeed, intrinsic randomness originates from latent random variables that are generally assimilated to noise when performing sensitivity analysis. So far, several approaches have been introduced.

The pragmatic approach for this purpose consists in performing GSA on both conditional expectation and conditional variance of the model output with respect to uncertain parameters. Both quantities are averaged quantities over the intrinsic randomness of the stochastic model. In practice, this comes down to estimate Sobol' indices (Sobol' [1993]) for two deterministic models. This approach is often used in practice in various applications, for instance in: [Courcou et al. \[2011\]](#) to identify key parameters of a model describing the spread of an animal disease in a cattle herd; [Rimbaud et al. \[2018\]](#) for a model describing the spatio-temporal spread of plant pathogens; [Richard et al. \[2021\]](#) for a SARS-CoV-2 spread model,

[Cristancho Fajardo et al. \[2021\]](#) for a theoretical metapopulation model. However, this approach can suffer from inconsistent conclusions. Since GSA is performed separately on conditional mean and conditional variance, a parameter can appear to be important for a quantity and not for the other one. One can check this on the toy example given by  $Y = X_1 + X_2Z$  with  $X_1, X_2, Z$  i.i.d. under standard normal distribution such that  $X_1, X_2$  are the inputs and  $Z$  stands for the intrinsic randomness variable. In this example, the first-order indices of  $X_1$  and  $X_2$  of the conditional expectation are respectively  $S_{X_1} = 1$  and  $S_{X_2} = 0$  whereas those of the conditional variance are  $S_{X_1} = 0$  and  $S_{X_2} = 1$ . Moreover, in [Mazo \[2021\]](#) it is shown that the Monte-Carlo estimator of first order Sobol' indices for conditional mean with respect to input is biased. So, estimation accuracy issues arise.

In addition, conditional expectations with respect to inputs require an averaging over the intrinsic randomness variable since the law of the latter is unknown. This induces a loss of information as this averaging affects inputs that interact with intrinsic randomness. In order to avoid this loss, [Hart et al. \[2017\]](#) introduced new sensitivity indices. For this purpose, Sobol'-Hoeffding decomposition is interpreted as a random decomposition, where the randomness is due to the intrinsic noise; corresponding Sobol' indices can also be considered as random. The sensitivity indices proposed by [Hart et al. \[2017\]](#) are then defined as expectations of the random Sobol' indices. However, they do not fully reflect interactions between intrinsic randomness and uncertain parameters.

Recently, new methods have been developed for stochastic models. They rely on a paradigm shift in the way of dealing with intrinsic stochasticity. The stochastic model is interpreted as a deterministic model with values in the set of probability distributions on the original output space. [Fort et al. \[2021\]](#) and [da Veiga \[2021\]](#) consider this approach and define new sensitivity indices well-suited to deterministic models with output valued in a set of probability distributions. This approach has the advantage of setting a framework in which stochastic output models are deterministic but it does not allow to assess the interactions between the intrinsic randomness and uncertain

parameters.

Furthermore, it should be noted that even though stochastic models are less studied than deterministic ones, not all types of stochastic models have received the same attention from the sensitivity analysis community. The most studied stochastic models are based on stochastic differential equation (see, e.g. [Le Maître and Knio \[2015\]](#), [Jimenez et al. \[2017\]](#), [Étoré et al. \[2020\]](#)). For these types of models, methods based on Polynomial Chaos Expansion meta-modeling have been proposed. Conversely, there are few GSA methods for models based on jump processes (Poisson processes, Markov chains, piecewise-deterministic jump processes). Generally, for these models, black-box or meta-model based GSA are proposed (e.g. [Marrel et al. \[2012\]](#), [Zhu and Sudret \[2021\]](#)).

In this paper, we focus on continuous-time Markov based models. We develop an approach for performing GSA for such models. This approach consists first in representing the model outputs as deterministic functions of a random vector of uncertain parameters and intrinsic randomness with known probability distribution and then in exploiting the resulting representation for GSA. This enables to put model output under a deterministic form so that contribution of uncertain parameters and intrinsic randomness as well as interaction of both can be assessed. Concretely, we study the continuous-time Markov chain given by the stochastic process that counts over time the number of individuals in each compartment of epidemic compartmental models. We rely on Gillespie Stochastic Simulation Algorithm ([Gillespie \[1976\]](#)) allowing exact simulation of Markov chains from which we derive a new representation. Furthermore, we include a second representation, the random time change introduced by Kurtz ([Ethier and Kurtz \[1986\]](#)) and studied in [Navarro Jimenez et al. \[2016\]](#) so as to achieve comparison of GSA results between the two representations. We apply the two approaches to a SARS-CoV-2 spread model.

The paper is organized as follows. In Section 2 we set the framework of stochastic models and associated representations. The definition of Sobol' indices is reminded and dependence of GSA results on the choice of the representation is discussed. Section 3 is dedicated to the description of representations of continuous-time Markov chains. For this purpose, we describe compartmental models and we discuss two different representations: Gillespie and Kurtz. In Section 4, we present the application of our approach to a SARS-CoV-2 spread model. This section includes a description of the model, the GSA results and comparison elements between the results obtained with the two different representations of the model introduced in Section 3.

## 2 The general approach

This section is devoted to introducing (Subsection 2.2) the concept of stochastic model representation under a deterministic form. Under such a form, GSA methods for deterministic models can be applied to perform sensitivity analysis. For this, the definition of the so-called Sobol' indices is briefly reminded in Subsection 2.1 in the context of deterministic models with scalar or functional outputs. In Subsection 2.3, we discuss the question of the dependence of GSA results to the choice of the representation of the stochastic model.

## 2.1 Global sensitivity analysis for deterministic functions

Let us consider a deterministic model  $g$  with input  $\mathbf{X} = (X_1, \dots, X_m)$  and output  $Y = g(\mathbf{X})$  so that  $\mathbb{E}[\|g(\mathbf{X})\|^2] < +\infty$  and  $X_1, \dots, X_m$  are mutually independent. For such a model, the definition of first-order and total Sobol' indices (Homma and Saltelli [1996]) is reminded according to two frameworks: a scalar output  $Y = g(\mathbf{X}) \in \mathbb{R}$  and a functional output  $g(\mathbf{X}) = (q(t, \mathbf{X}); t \in \{t_0, \dots, T\})$  for which dynamical and aggregated Sobol' indices are introduced.

In scalar case, first-order and total Sobol' indices (Sobol' [1993]) associated to each input  $X_j, j = 1, \dots, m$  are defined as:

$$S_{X_j} = \frac{\text{Var}(\mathbb{E}[g(\mathbf{X}) | X_j])}{\text{Var}(g(\mathbf{X}))},$$

$$S_{X_j}^T = 1 - \frac{\text{Var}(\mathbb{E}[g(\mathbf{X}) | X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_m])}{\text{Var}(g(\mathbf{X}))}.$$

In the functional case, we focus on two types of indices: dynamical Sobol' indices and aggregated Sobol' indices. Considering each random variable  $q(t, \mathbf{X})$ , sensitivity indices such as first-order Sobol' indices and total Sobol' indices can be obtained as in scalar case for each time  $t$  as  $\mathbb{E}[q(t, \mathbf{X})^2]$  is finite. Regarding aggregated Sobol' indices, note that output  $g(\mathbf{X})$  is multidimensional so that according to Lamboni et al. [2011] and Gamboa et al. [2014], first-order and total Sobol' indices associated to each input  $X_j, j = 1, \dots, m$  are respectively given by:

$$GSI_{X_j} = \frac{\text{Trace}[\text{Var}(\mathbb{E}[g(\mathbf{X}) | X_j])]}{\text{Trace}[\text{Var}(g(\mathbf{X}))]},$$

$$GSI_{X_j}^T = 1 - \frac{\text{Trace}[\text{Var}(\mathbb{E}[g(\mathbf{X}) | X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_m])]}{\text{Trace}[\text{Var}(g(\mathbf{X}))]},$$

where  $\text{Var}(\mathbb{E}[g(\mathbf{X}) | X_j])$  and  $\text{Var}(g(\mathbf{X}))$  are the variance-covariance matrices of the random vectors  $\mathbb{E}[g(\mathbf{X}) | X_j]$  and  $g(\mathbf{X})$ , respectively.

## 2.2 Deterministic representations of stochastic models

In the following, a stochastic model  $g$  is defined as a random function taking  $\theta \in \Theta \subset \mathbb{R}^p$  as input and producing an output  $g(\theta)$  which is a random variable with values in a set  $\mathcal{Y}$ . If  $f : \Theta \times \mathcal{Z} \rightarrow \mathcal{Y}$  is a deterministic function and  $Z$  is a random element valued in a set  $\mathcal{Z}$  such that  $(\mathbf{X}, g(\mathbf{X}))$  and  $(\mathbf{X}, f(\mathbf{X}, Z))$  are identically distributed, then the couple  $(f, Z)$  is said to be a deterministic representation of the stochastic model  $g$ . An example of stochastic model with two representations is provided in Example 1.

**Example 1** (Toy example). *Let  $U \sim \mathcal{U}([0, 1])$  independent of  $(\mathbf{X}, Z, Z_1, Z_2) \sim \mathcal{N}(0_{\mathbb{R}^4}, Id_4)$ . Consider the stochastic model  $g$  with input  $\mathbf{X}$  and output  $g(\mathbf{X}) = \mathbf{X} + Z$ . This model can be represented by using  $f(\mathbf{X}, Z_1, Z_2) = \mathbf{X} + \frac{1}{\sqrt{2}}(Z_1 + Z_2)$  or  $f'(\mathbf{X}, U) = \mathbf{X} + \Phi^{-1}(U)$ , where  $\Phi$  is the cumulative distribution function of the standard normal distribution.*

The function  $f$  is not necessarily explicit. It can correspond to an algorithm. Proposition 1 (see Appendix B for the proof) provides a sufficient condition for a couple  $(f, Z)$  to be a representation of a stochastic model  $g$ .

**Proposition 1.** *Assume that  $\mathbf{X}$  and  $Z$  are independent and that for all  $\theta \in \Theta$ , the probability distributions of  $g(\theta)$  and  $f(\theta, Z)$  are identical. Then  $(f, Z)$  is a deterministic representation of the stochastic model  $g$ .*

Representing stochastic models under deterministic form is useful in GSA. If both the probability distribution of  $Z$  and the function  $f$  are known, then the stochastic model becomes a deterministic model with inputs  $(\mathbf{X}, Z)$ . Hence, all the standard methods of GSA such as those presented in Subsection 2.1 can be applied. Therefore, contribution of  $Z$  and its interaction with uncertain parameters can be assessed.

However, as already discussed (see Example 1), a stochastic model may admit several representations. Indeed, the way to simulate a random variable is not unique, leading to different stochastic simulators. Considering two distinct representations  $(f, Z)$  and  $(f', Z')$  of the same stochastic model necessarily yields that the joint probability distributions of  $(\mathbf{X}, f(\mathbf{X}, Z))$  and  $(\mathbf{X}, f'(\mathbf{X}, Z'))$  are identical. However, the intrinsic randomness elements  $Z$  and  $Z'$  and functions  $f$  and  $f'$  may differ from one representation to the other. A natural question is then whether GSA results depend on the chosen representation. The answer is yes; this point is discussed in Subsection 2.3.

### 2.3 Dependence of global sensitivity analysis on model representation

In Subsection 2.2, it appears that representation always preserves the probability distribution of  $(\mathbf{X}, g(\mathbf{X}))$ . Our aim in this section is to prove that the results of GSA depends on the choice of representation.

Let us consider  $(f, Z)$  and  $(f', Z')$  two distinct representations of the stochastic model  $g$  with input  $\mathbf{X} = (X_1, \dots, X_m)$  and output  $g(\mathbf{X})$ . By definition,  $(\mathbf{X}, g(\mathbf{X})) \sim (\mathbf{X}, f(\mathbf{X}, Z)) \sim (\mathbf{X}, f'(\mathbf{X}, Z'))$ . If  $u$  is a subset of  $\{1, \dots, m\}$ , then:  $\mathbb{E}[f(\mathbf{X}, Z) | \mathbf{X}_u] = \mathbb{E}[f'(\mathbf{X}, Z') | \mathbf{X}_u]$  almost surely with  $\mathbf{X}_u = \{X_j, j \in u\}$  (see Mazo [2021]). So, it yields:

$$\frac{\text{Var}(\mathbb{E}[f(\mathbf{X}, Z) | \mathbf{X}_u])}{\text{Var}(f(\mathbf{X}, Z))} = \frac{\text{Var}(\mathbb{E}[f'(\mathbf{X}, Z') | \mathbf{X}_u])}{\text{Var}(f'(\mathbf{X}, Z'))}.$$

It implies that closed Sobol' indices associated to  $\mathbf{X}_u, u \subseteq \{1, \dots, m\}$  are representation-free, i.e. they do not depend on the chosen representation. A straightforward consequence is the invariance of first-order Sobol' indices associated to each  $X_j, j = 1, \dots, m$ , and of the total Sobol' index associated to  $Z$  with respect to the choice of the representation.

However,  $\mathbb{E}[f(\mathbf{X}, Z) | (\mathbf{X}_u, Z)]$  and  $\mathbb{E}[f'(\mathbf{X}, Z') | (\mathbf{X}_u, Z')]$  may have different probability distributions. Indeed, since  $(f, Z) \neq (f', Z')$ , the way each function  $f$  or  $f'$  combines its relative intrinsic randomness variable with input  $\mathbf{X}$  to generate outputs is different. So, differences can appear on some quantities such as conditional expectations with respect to a group of random variables that includes the intrinsic randomness variable. This is illustrated in Example 2.

**Example 2.** *Let  $X$  be a random variable independent of  $Z$  and  $Z'$  where  $Z$  and  $Z'$  are i.i.d. under  $\mathcal{N}(0, 1)$ . Define two functions:  $f(X, Z) = XZ$  and  $f'(X, Z') = X^2Z'$ . If  $X$  is distributed such that  $\mathbb{P}(X = -1) = \mathbb{P}(X = 1) = \frac{1}{2}$  then  $(X, f(X, Z)) \sim (X, f'(X, Z'))$ . So  $(f, Z)$  and  $(f', Z')$  represent the same stochastic model but:*

$$\mathbb{E}[f(X, Z) | Z] = 0 \quad \mathbb{E}[f'(X, Z') | Z'] = Z'.$$

First-order Sobol' indices of intrinsic randomness in the two representations are given by:

$$S_Z(f) = \frac{\text{Var}(\mathbb{E}[f(X, Z) | Z])}{\text{Var}(f(X, Z))} = 0 \text{ while } S_Z(f') = \frac{\text{Var}(\mathbb{E}[f'(X, Z') | Z'])}{\text{Var}(f'(X, Z))} = 1.$$

And total Sobol' indices of  $X$  in the two representations are different:  $S_X^T(f) = 1 - S_Z(f) = 1$  and  $S_X^T(f') = 1 - S_Z(f') = 0$ . Consequently, first-order Sobol' index associated to intrinsic randomness depends on the choice of the representation.

Via a counter-example (see Example 2) and theoretical elements, we showed that GSA results depend on the choice of the representation. This point is illustrated in Section 4 in the context of GSA of continuous-time Markov chain based models, which are the models of interest in this paper.

### 3 Deterministic representations of CTMC stochastic compartmental models

In this section, we discuss two different representations for CTMC stochastic compartmental models. The section is organized as follows. In Subsection 3.1, we provide a description of the CTMC under study by using graph formalism to represent compartmental models. In Subsection 3.2, we present the representation detailed in Navarro Jimenez et al. [2016] and then we introduce a new representation based on Gillespie Stochastic Simulation Algorithm (SSA).

#### 3.1 CTMC stochastic compartmental models

Consider a closed population that includes  $N$  individuals (i.e.  $N$  remains constant over time). Assume that an epidemic outbreaks within this population. Individuals can be susceptible or at various stages of infections. So, at each time, each individual is in a certain health status. Then, individuals are grouped according to their health status. The resulting groups form a partition (compartment). Let  $\mathbf{V}$  be the set of compartments. As health status of each individual can change over time, transitions can take place between compartments. A transition always involves two different compartments, say  $\alpha, \beta \in \mathbf{V}$  such that  $\alpha \neq \beta$ . A pair of compartments  $(\alpha, \beta)$  between which transitions are possible defines a type of transition occurring in the direction  $\alpha \rightarrow \beta$ . So, the pair  $(\alpha, \beta)$  forms an arrow. An individual can move from a compartment  $\alpha$  to another  $\beta$  only if there is an arrow from  $\alpha$  to  $\beta$ . To each arrow is associated a vector  $\mathbf{u}_{\alpha, \beta} \in \{-1, 0, +1\}^{\text{card } \mathbf{V}}$  whose components are zero except at the components corresponding to  $\alpha$  and  $\beta$  which are equal to  $-1$  and  $+1$  respectively. Denote  $\mathbf{E}$  the set of arrows and  $n_E$  its cardinal. The couple  $\mathcal{G} = (\mathbf{V}, \mathbf{E})$  is a directed graph with vertices  $\mathbf{V}$  and edges  $\mathbf{E}$ .

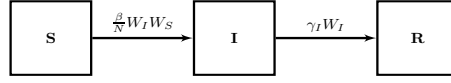
The intensities of transitions between compartments depend on the specific parameters  $\theta$  of the epidemic. As  $\theta$  is generally unknown, let  $\Theta \subset \mathbb{R}^d$  be the set of all possible parameters  $\theta$ . We are interested in the dynamics of the number of individuals in each compartment over time, which are stochastic since transitions occur at random times. Let  $W_\alpha^\theta(t)$  denote the number of individuals in compartment  $\alpha$  at time  $t$  and suppose that the initial state  $\xi_0 \in \{0, \dots, N\}^{\text{card } \mathbf{V}}$  of the process  $W^\theta := \{(W_\alpha^\theta(t))_{\alpha \in \mathbf{V}}; t \geq 0\}$  is known. Assume that for each  $\theta \in \Theta$ , the stochastic process  $W^\theta$  is a homogeneous continuous-time Markov chain with state space  $\mathcal{E}$  defined as the the smallest subset of  $\{0, \dots, N\}^{\text{card } \mathbf{V}}$  that contains all the vectors of the

form  $\xi_0 + \sum_{i=1}^n \mathbf{u}_{\alpha^{(i)}, \beta^{(i)}}$  where  $n \in \mathbb{N}$  and  $(\alpha^{(i)}, \beta^{(i)}) \in \mathbf{E}$  for all  $i = 1, \dots, n$ . The generator  $\mathcal{Q}_\theta$  of  $W^\theta$  is characterized by nonnegative rate functions  $g_{\alpha, \beta}$  defined by:  $g_{\alpha, \beta}(\theta, \xi) = \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \mathbb{P}(W^\theta(s + \varepsilon) = \xi + \mathbf{u}_{\alpha, \beta} \mid W^\theta(s) = \xi)$  for  $\xi \in \mathcal{E}$  and  $s \geq 0$  so that the element of the generator at the row corresponding to  $\xi \in \mathcal{E}$  and column corresponding to  $\xi' \in \mathcal{E}$  is given by:

$$\mathcal{Q}_\theta(\xi, \xi') = \begin{cases} g_{\alpha, \beta}(\theta, \xi) & \text{if } \xi' = \xi + \mathbf{u}_{\alpha, \beta} \\ -\sum_{(\alpha', \beta') \in \mathbf{E}} g_{\alpha', \beta'}(\theta, \xi) & \text{if } \xi' = \xi \\ 0 & \text{otherwise.} \end{cases}$$

Every transition  $\xi \rightarrow \xi + \mathbf{u}_{\alpha, \beta}$  of type  $(\alpha, \beta) \in \mathbf{E}$  occurs at rate  $g_{\alpha, \beta}(\theta, \xi)$ .

**Example 3.** Let us consider the classical SIR model:



There are three compartments  $\mathbf{V} = \{S, I, R\}$  and two types of transitions: infection  $(S, I)$  and removal  $(I, R)$  so that  $\mathbf{E} = \{(S, I), (I, R)\}$ . Infection is characterized by transition vector  $\mathbf{u}_{S, I} = (-1, +1, 0)$  and rate function  $g_{S, I} = \frac{\beta}{N} W_I W_S$ . Removal has transition vector  $\mathbf{u}_{I, R} = (0, -1, +1)$  and rate function  $g_{I, R} = \gamma_I W_I$ .

Commonly, Gillespie direct method is used to simulate continuous-time Markov chains. Algorithm 1 provides instructions intended to simulating paths of the process  $W^\theta$  for a given  $\theta \in \Theta$ .

---

**Algorithm 1:** Gillespie SSA (direct method)

---

**Data:**  $\xi_0, \{g_{\alpha^{(j)}, \beta^{(j)}}, j = 1, \dots, n_E\}, \{\mathbf{u}_{\alpha^{(j)}, \beta^{(j)}}, j = 1, \dots, n_E\}, T$   
**inputs:**  $\theta$   
**output:**  $W = \{W(t); t \in [0, T]\}$   
 initialization;  
 $t \leftarrow 0, W(t) \leftarrow \xi_0;$   
**while**  $t < T$  **do**  
    $\lambda \leftarrow \sum_{j=1}^{n_E} g_{\alpha^{(j)}, \beta^{(j)}}(\theta, W(t));$   
   Draw  $\tau$  with exponential distribution with mean  $1/\lambda;$   
   **for**  $j = 1$  **to**  $n_E$  **do**  
      $p_j \leftarrow g_{\alpha^{(j)}, \beta^{(j)}}(W(t), \theta) / \lambda;$   
   **end**  
   Draw  $U$  with standard uniform distribution;  
   Pick  $l \in \{1, \dots, n_E\}$  with distribution  $(p_1, \dots, p_{n_E})$  using  $U;$   
    $W(t + \tau) \leftarrow W(t) + \mathbf{u}_{\alpha^{(l)}, \beta^{(l)}};$   
    $t \leftarrow t + \tau;$   
**end**

---

### 3.2 Deterministic representations

Assume that the random vector  $\mathbf{X}$  models the parameter uncertainty. Then, consider the stochastic model  $(\mathbf{X}, W^{\mathbf{X}})$ . For simplicity, let denote  $W^{\mathbf{X}}$  by  $\mathbf{W}$ . We



seek deterministic representations  $(f, Z)$  of  $(\mathbf{X}, \mathbf{W})$ , i.e. such that the probability distribution of  $Z$  is known. For this aim, thanks to Proposition 1, it is sufficient to find  $Z$  independent of  $\mathbf{X}$  and  $f$  such that: for all  $\theta \in \Theta$ ,  $W^\theta \sim f(\cdot, \theta, Z) = \{f(t, \theta, Z); t \in [0, T]\}$ . In the following, we discuss two representations of  $(\mathbf{X}, \mathbf{W})$ . We present the random-time change representation studied in [Le Maître et al. \[2015\]](#) and [Navarro Jimenez et al. \[2016\]](#) and introduce the new representation based on Gillespie algorithm.

### 3.2.1 Random time change representation

The random time change representation is based on the random time change decomposition of the process  $W^\theta$  for each  $\theta \in \Theta$ . This decomposition has been introduced by [Ethier and Kurtz \[1986\]](#). Consider the vector  $Z^K = (Z_{\alpha, \beta})_{(\alpha, \beta) \in \mathbf{E}}$  of independent unit-rate Poisson processes  $Z_{\alpha, \beta}(\cdot)$ . [Kurtz \[1982\]](#) showed that there exists a function  $f_K$  satisfying

$$f_K(t, \theta, Z^K) = \xi_0 + \sum_{(\alpha, \beta) \in \mathbf{E}} Z_{\alpha, \beta} \left( \int_0^t g_{\alpha, \beta}(\theta, f_K(s, \theta, z^K)) ds \right) \mathbf{u}_{\alpha, \beta},$$

which defines a continuous-time Markov chain with initial state  $\xi_0$  and generator  $\mathcal{Q}_\theta$ . So,  $f_K(\cdot, \theta, Z^K) \sim W^\theta$  for all  $\theta \in \Theta$ . In addition,  $Z^K$  does not depend on  $\Theta$ . By construction,  $\mathbf{X}$  and  $Z^K$  are independent. Hence  $(f_K, Z^K)$  defines a representation of  $(\mathbf{X}, \mathbf{W})$ .

For each  $(\alpha, \beta) \in \mathbf{E}$ ,  $Z_{\alpha, \beta}$  stands for the intrinsic noise of the reaction or type of transition  $(\alpha, \beta)$ . So,  $Z^K$  includes intrinsic noise of each reaction channel or type of transition. Since the Poisson processes  $Z_{\alpha, \beta}, (\alpha, \beta) \in \mathbf{E}$  that compose  $Z^K$  are mutually independent, then this representation enables to assess contribution of intrinsic noise of each reaction channel or type of transition as well as that of the whole intrinsic randomness  $Z^K$ . [Le Maître et al. \[2015\]](#) and [Navarro Jimenez et al. \[2016\]](#) used this representation to perform GSA for chemical reaction network models in order to estimate contribution of uncertain parameters and reaction channel intrinsic noises. In practice, simulation of this representation relies on the Modified Next Reaction Method (MNRM) developed by [Anderson \[2007\]](#) and provided in Algorithm 2 ([Navarro Jimenez et al. \[2016\]](#)).

---

**Algorithm 2: Modified Next Reaction Method**

---

**Data:**  $\xi_0, \{g_{\alpha^{(j)}, \beta^{(j)}}, j = 1, \dots, n_E\}, \{\mathbf{u}_{\alpha^{(j)}, \beta^{(j)}}, j = 1, \dots, n_E\}, T$   
**inputs:**  $\theta, Z = (RG_1, \dots, RG_{n_E})$   
**output:**  $W = \{W(t), t \in [0, T]\}$   
initialization;  
**for**  $j = 1$  **to**  $n_E$  **do**  
    | Draw  $r_j$  from  $RG_j$ ;  
    |  $t_j \leftarrow 0, t_j^+ \leftarrow -\log(r_j)$ ;  
**end**  
 $t \leftarrow 0, W(t) \leftarrow \xi_0$ ;  
**while**  $t < T$  **do**  
    | **for**  $j = 1$  **to**  $n_E$  **do**  
        | Evaluate  $a_j = g_{\alpha^{(j)}, \beta^{(j)}}(\theta, W(t))$  and  $dt_j = \frac{t_j^+ - t_j}{a_j}$   
        **end**  
        Set  $s = \operatorname{argmin}_j dt_j$ ;  
         $W(t + dt_s) \leftarrow W(t) + \mathbf{u}_{\alpha^{(s)}, \beta^{(s)}}$ ;  
         $t \leftarrow t + dt_s$ ;  
        **for**  $j = 1$  **to**  $n_E$  **do**  
            |  $t_j \leftarrow t_j + a_j \cdot dt_s$   
        **end**  
        Get  $r_s$  from  $RG_s$  and set  $t_s^+ \leftarrow t_s^+ - \log(r_s)$   
    **end**

---

### 3.2.2 Gillespie representation

Gillespie SSA is widely used to simulate Markov chains. Especially, in epidemiology, many CTMC model simulators are based on this algorithm. Both Gillespie SSA and the Modified Next Reaction Method generate statistically exact paths of CTMC including  $W^\theta$ : they are stochastically equivalent. However, representations of  $(\mathbf{X}, \mathbf{W})$  derived from these two algorithms can lead to different sensitivity analysis results corresponding to different information as the interpretation of intrinsic noise is not the same for both representations. In the following, we introduce a representation of  $(\mathbf{X}, \mathbf{W})$  based on Gillespie SSA.

Gillespie SSA simulates two processes: the jump process that provides the jump times of  $W^\theta$  and the embedded chain that describes successive states of  $W^\theta$ . Given two independent sequences of i.i.d. standard uniform variables  $\mathbf{U}^1 = (U_1^1, U_2^1, \dots)$  and  $\mathbf{U}^2 = (U_1^2, U_2^2, \dots)$ , assume that the jump process is simulated from  $\mathbf{U}^1$  and the embedded chain from  $\mathbf{U}^2$ . In practice,  $\mathbf{U}^1$  and  $\mathbf{U}^2$  are respectively assimilated to two numbers  $RG_1$  and  $RG_2$  which are used as a seed for the random number generator so that setting seed to a value enables to stream random numbers. The resulting algorithm that is detailed in Algorithm 3 is a modification of the Gillespie SSA provided in Algorithm 1:

---

**Algorithm 3:**

---

**Data:**  $\xi_0, \{g_{\alpha^{(j)}, \beta^{(j)}}, j = 1, \dots, n_E\}, \{\mathbf{u}_{\alpha^{(j)}, \beta^{(j)}}, j = 1, \dots, n_E\}, T$   
**inputs:**  $\theta, Z = (RG_1, RG_2)$   
**output:**  $W = \{W(t); t \in [0, T]\}$   
initialization;  
 $t \leftarrow 0, W(t) \leftarrow \xi_0, i \leftarrow 1;$   
**while**  $t < T$  **do**  
    **for**  $k = 1$  **to**  $2$  **do**  
        | Pick  $U_i^k$  from  $RG_k$ ;  
    **end**  
     $\lambda \leftarrow \sum_{j=1}^{n_E} g_{\alpha^{(j)}, \beta^{(j)}}(\theta, W(t));$   
     $t^* \leftarrow -\log(U_i^1)/\lambda;$   
    **for**  $j = 1$  **to**  $n_E$  **do**  
        |  $p_j \leftarrow g_{\alpha^{(j)}, \beta^{(j)}}(W(t), \theta)/\lambda;$   
    **end**  
    Pick  $l$  such that  $\sum_{j=1}^{l-1} p_j \leq U_i^2 < \sum_{j=1}^l p_j;$   
     $W(t + t^*) \leftarrow W(t) + \mathbf{u}_{\alpha^{(l)}, \beta^{(l)}};$   
     $t \leftarrow t + t^*;$   
     $i \leftarrow i + 1;$   
**end**

---

Let  $Z^G$  be the vector  $(\mathbf{U}^1, \mathbf{U}^2)$ . If  $f_G(t, \theta, Z^G)$  denotes the output of Algorithm 3 at time  $t$  for input  $(\theta, Z^G)$ , then the following result holds:

**Proposition 2.** Denote  $\mathcal{L}([0, 1])$  the space of sequences on  $[0, 1]$ . The function:

$$f_G : \mathbb{R}_+ \times \Theta \times \mathcal{L}([0, 1])^2 \longrightarrow \mathcal{E}$$
$$(t, \theta, z^G) \longmapsto f_G(t, \theta, z^G)$$

is such that the continuous-time Markov chains  $W^\theta$  and  $f_G(\cdot, \theta, Z^G) := \{f_G(t, \theta, Z^G); t \geq 0\}$  have the same finite-dimensional distributions.

The proof of Proposition 2 is detailed in Appendix B. Note that  $Z^G$  does not depend on  $\theta$  so that  $Z^G$  and  $\mathbf{X}$  are independent. Therefore, Propositions 1 and 2 ensure that  $(f_G, Z^G)$  is a representation of  $(\mathbf{X}, \mathbf{W})$ . The random vector  $Z^G = (\mathbf{U}^1, \mathbf{U}^2)$  stands for the intrinsic randomness variable.  $Z^G$  includes two intrinsic noises  $\mathbf{U}^1$  and  $\mathbf{U}^2$  that correspond to intrinsic noises of the jump time process and the embedded discrete chain.

Algorithm 3 aggregates all the types of transition processes to generate the sequence of jump times and the discrete chain. So, intrinsic noise associated to each type of transition  $(\alpha, \beta) \in \mathbf{E}$  cannot be identified with this algorithm. Therefore, it is not possible to assess contribution of intrinsic noises of type of transition processes. To overcome this insufficiency, we can rely on the first reaction method studied by Gillespie (1976) to build a representation allowing to separate the intrinsic noises associated to each type of transition processes.

For this purpose, let  $Z^G$  be the random vector  $(\mathbf{U}_{\alpha, \beta}, (\alpha, \beta) \in \mathbf{E})$  where each  $\mathbf{U}_{\alpha, \beta}$  is a sequence of i.i.d. standard uniform variables and assume that each component of  $Z^G$  is identified to a number  $RG_j$  with  $j = 1, \dots, n_E$ . A modification of Gillespie

first reaction method algorithm in a similar way as Gillespie direct method yields Algorithm 4:

---

**Algorithm 4:**

---

**Data:**  $\xi_0, \{g_{\alpha^{(j)}, \beta^{(j)}}, j = 1, \dots, n_E\}, \{\mathbf{u}_{\alpha^{(j)}, \beta^{(j)}}, j = 1, \dots, n_E\}, T$   
**inputs:**  $\theta, Z = (RG_1, \dots, RG_{n_E})$   
**output:**  $W = \{W(t); t \in [0, T]\}$   
initialization;  
 $t \leftarrow 0, W(t) \leftarrow \xi_0, i \leftarrow 1;$   
**while**  $t < T$  **do**  
    **for**  $j = 1$  **to**  $n_E$  **do**  
        Pick  $U_{\alpha^{(j)}, \beta^{(j)}}^i$  from  $RG_j$ ;  
        Evaluate  $dt_j = \frac{-\log(U_{\alpha^{(j)}, \beta^{(j)}}^i)}{g_{\alpha^{(j)}, \beta^{(j)}}(\theta, W(t))}$   
    **end**  
    Set  $l = \operatorname{argmin}_j dt_j$ ;  
     $W(t + dt_l) \leftarrow W(t) + \mathbf{u}_{\alpha^{(l)}, \beta^{(l)}}$ ;  
     $t \leftarrow t + dt_l$ ;  
     $i \leftarrow i + 1$   
**end**

---

If  $f_G(t, \theta, Z^G)$  denotes the output of Algorithm 4 at time  $t$  and input  $(\theta, Z^G)$ , it yields:

**Proposition 3.** Denote  $\mathcal{L}([0, 1])$  the space of sequences on  $[0, 1]$ . The function:

$$f_G : \mathbb{R}_+ \times \Theta \times \mathcal{L}([0, 1])^{n_E} \longrightarrow \mathcal{E}$$

$$(t, \theta, z^G) \longmapsto f_G(t, \theta, z^G)$$

is such that the continuous-time Markov chains  $W^\theta$  and  $f_G(\cdot, \theta, Z^G) := \{f_G(t, \theta, Z^G); t \geq 0\}$  have the same finite-dimensional distributions.

Proposition 3 provides a new representation of  $(\mathbf{X}, \mathbf{W})$  since  $\mathbf{X}$  and  $Z^G$  are independent. This representation allows to separate intrinsic randomness into independent intrinsic noises of type of transition processes. The proof of Proposition 3 is detailed in Appendix B.

The intrinsic randomness captures all the variability of the model which does not depend on epidemic parameters. From an epidemiological point of view, it encompasses the random behavior of individuals and their social interactions, as well as the biological variability between individuals with respect to infection.

## 4 Application to an epidemic model

In this section, we implement our method on a parsimonious SARS-CoV-2 spread model inspired by the literature (Cazelles et al. [2021]). We perform variance-based GSA using the two representations: the one analyzed in Navarro Jimenez et al. [2016] and based on the random time change (described in Section 3.2.1) and the new one we introduced in Section 3.2.2 based on the modification of Gillespie algorithm provided in Proposition 3. The model considered in this section, although not necessarily the

most refined from the point of view of application, remains very relevant. Moreover, it should be noted that our approach is generic and can therefore be applied to any other compartmental model.

#### 4.1 Description of the model of SARS-CoV-2 spread

The compartmental model in Figure 1 is used to describe the spread of SARS-CoV-2 within a closed population with size  $N$ . This model includes seven compartments corresponding to seven health states and nine transitions between these states. Indeed, in this model, an individual can be susceptible (S), exposed (E) (i.e. infected but not yet infectious), asymptomatic infectious (A), symptomatic infectious (I), hospitalized (H), recovered (R) or dead (D). Two modeling assumptions can be mentioned. First, infection is neglected within hospitals so that hospitalized individuals cannot infect. Secondly, it is assumed that recovered individuals get perfectly immunized, so they cannot be susceptible after recovering. This assumption is valid on short time intervals.

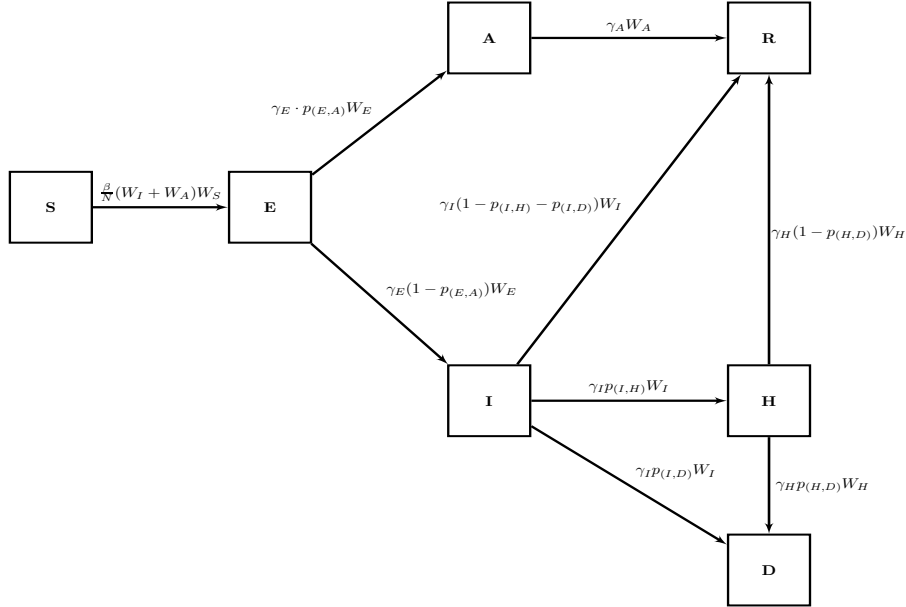


Figure 1: Compartmental model of the spread of SARS-CoV-2 within a population, comprising seven health states (S,E,A,I,H,R,D) and corresponding transition rates.  $W_\alpha$  with  $\alpha \in \{S, E, A, I, H, R, D\}$  denotes the number of individuals in compartment  $\alpha$ .

We are interested in the process  $\{(W_S(t), W_E(t), W_A(t), W_I(t), W_H(t), W_R(t), W_D(t)); t \in [0, T]\}$  that takes values in  $\mathcal{E} \subset \{0, \dots, N\}^7$  and counts over time the number of individuals in each compartment, where  $N$  is the size of the population and  $T \in (0, +\infty]$  is the final time of the study. The set of compartments is given by  $\mathbf{V} = \{S, E, A, I, H, R, D\}$ . The different types of transitions and their characteristics (rate function  $g_{\alpha, \alpha'}$  and transition vector  $\mathbf{u}_{\alpha, \alpha'}$ , where  $(\alpha, \alpha')$  denotes a type of transition of the model) are described in Table 1.

Type of transition ( $\alpha, \alpha'$ )	Type	Rate function ( $g_{\alpha, \alpha'}$ )	Transition vector ( $\mathbf{u}_{\alpha, \alpha'}$ )
$(S, E)$	infection	$\frac{\beta}{N} \cdot W_S \cdot (W_A + W_I)$	$(-1, 1, 0, 0, 0, 0, 0)$
$(E, A)$	asymptomatic infection activation	$\gamma_E \cdot p_{(E,A)} \cdot W_E$	$(0, -1, +1, 0, 0, 0, 0)$
$(E, I)$	symptomatic infection activation	$\gamma_E \cdot (1 - p_{(E,A)}) \cdot W_E$	$(0, -1, 0, +1, 0, 0, 0)$
$(A, R)$	recovery of an asymptomatic	$\gamma_A \cdot W_A$	$(0, 0, -1, 0, 0, +1, 0)$
$(I, R)$	recovery of a symptomatic	$\gamma_I \cdot (1 - p_{(I,H)} - p_{(I,D)}) \cdot W_I$	$(0, 0, 0, -1, 0, +1, 0)$
$(I, H)$	hospitalization of a symptomatic	$\gamma_I \cdot p_{(I,H)} \cdot W_I$	$(0, 0, 0, -1, +1, 0, 0)$
$(I, D)$	death of a symptomatic	$\gamma_I \cdot p_{(I,D)} \cdot W_I$	$(0, 0, 0, -1, 0, 0, +1)$
$(H, R)$	recovery of a hospitalized	$\gamma_H \cdot (1 - p_{(H,D)}) \cdot W_H$	$(0, 0, 0, 0, -1, +1, 0)$
$(H, D)$	death of a hospitalized	$\gamma_H \cdot p_{(H,D)} \cdot W_H$	$(0, 0, 0, 0, -1, 0, +1)$

Table 1: Description of the model transitions between states  $\{S, E, A, I, H, R, D\}$ .

Rate functions depend on the vector parameter  $\theta = (\beta, \gamma_E, \gamma_A, \gamma_I, \gamma_H, p_{(E,A)}, p_{(I,H)}, p_{(I,D)}, p_{(H,D)})$ . The interpretation of each of these parameters and their ranges of variation in the sensitivity analysis are provided in Table 2.

Parameter Name	Parameter Role	Nominal value	Range of variation
$\beta$	transmission rate	2	$(0.35, 4)$
$1/\gamma_E$	mean sojourn duration in $E$	4.6 days	$(2, 7)$
$1/\gamma_A$	mean sojourn duration in $A$	2.1 days	$(1, 3)$
$1/\gamma_I$	mean sojourn duration in $I$	4 days	$(3, 5)$
$1/\gamma_H$	mean sojourn duration in $H$	10 days	$(7, 12)$
$p_{(E,A)}$	probability for an exposed to become asymptomatic	0.6	$(0.3, 0.7)$
$p_{(I,H)}$	probability for a symptomatic to be hospitalized	0.15	$(10^{-3}, 0.2)$
$p_{(I,D)}$	probability for a symptomatic to die	0.05	$(10^{-3}, 0.1)$
$p_{(H,D)}$	probability for a hospitalized to die	0.08	$(0.001, 0.1)$

Table 2: Model parameter nominal values and their range of variation in the sensitivity analysis (values are plausible with current knowledge e.g. [Knock et al. \[2021\]](#)).

The process under study is:  $W^\theta = \{(W_\alpha(t))_{\alpha \in \mathbf{V}}; t \in [0, T]\}$ . Let  $\Theta$  be the set of all possible values of  $\theta$ . For all  $\theta \in \Theta$ ,  $W^\theta$  is assumed to be a continuous-time Markov chain.

Let denote  $p_I$  the vector  $(p_{(I,H)}, p_{(I,D)})$ . For GSA purposes, we focus on the group of inputs  $p_I$  instead of  $p_{(I,H)}$  and  $p_{(I,D)}$  separately for two reasons. The first reason is related to the assumption of mutual independence of inputs that is necessary for Sobol'-

Hoeffding decomposition. Since  $p_{(I,H)}$  and  $p_{(I,D)}$  are correlated, they are grouped as one input in order to ensure input independence. The second reason is that we are interested in the global influence of the probability distribution  $(p_{(I,H)}, p_{(I,D)}, 1 - p_{(I,H)} - p_{(I,D)})$  that is represented by  $p_I$ . Therefore, in the context of sensitivity analysis, there are eight parameter inputs considered:  $\beta, \gamma_E, \gamma_A, \gamma_I, \gamma_H, p_{(E,A)}, p_I, p_{(H,D)}$ . Let us model the uncertainty by introducing the random vector  $\mathbf{X} = (\beta, \gamma_E, \gamma_A, \gamma_I, \gamma_H, \mathbf{P}_{(E,A)}, \mathbf{P}_I, \mathbf{P}_{(H,D)})$  such that components are mutually independent with uniform distributions on intervals specified in Table 2. Specifically, the vector input  $\mathbf{P}_I$  is drawn from a multidimensional uniform on  $[10^{-3}, 0.2] \times [10^{-3}, 0.1]$ .

We study two types of outputs that are our quantities of interest (QoIs): a scalar QoI and a functional one. The scalar QoI considered is the extinction time  $Y_{ext}$  of the epidemic.  $Y_{ext}$  is defined as the first time when there are no more individuals  $E$  able to get infected or individuals  $A$  or  $I$  able to maintain infection by infecting new susceptible individuals:

$$Y_{ext} = \inf\{t \geq 0 : W_E(t) + W_A(t) + W_I(t) = 0\}.$$

Note that for all  $\theta \in \Theta$ ,  $Y_{ext}$  is well-defined, i.e.  $Y_{ext} < +\infty$ . Indeed, by considering the compartmental model described in Figure 1, after a finite number of transitions, the stochastic process will necessarily reach an absorbing state where there is no individual in compartments  $E, A$  and  $I$ . A boxplot of  $n = 2000$  simulations of  $Y_{ext}$  with parameter inputs set to nominal values given in Table 2 is showed in Figure S1.

The functional QoI is given by the dynamic of the number of symptomatic infectious individuals:

$$Y_I = \{W_I(t), t \in [0, T]\}.$$

The quantities  $Y_{ext}$  and  $Y_I$  are functions of the random field  $\mathbf{W} = \{W^{\mathbf{X}}(t), t \in [0, T]\}$ , so they inherit its representations.

By relying on Subsection 3.2, we obtained two representations of  $\mathbf{W}$ : the Gillespie representation  $(\mathbf{X}, \mathbf{W}) \sim (\mathbf{X}, f_G(\mathbf{X}, Z^G))$  where  $Z^G = (\mathbf{U}^1, \dots, \mathbf{U}^9)$  is a vector of nine independent sequences (one for each type of transition) of i.i.d. standard uniform variables and the Random time change representation  $(\mathbf{X}, \mathbf{W}) \sim (\mathbf{X}, f_K(\mathbf{X}, Z^K))$  where  $Z^K = (Z^1, \dots, Z^9)$  is a vector of nine independent unit-rate Poisson processes. For sensitivity analysis, intrinsic randomness variables  $Z^G$  and  $Z^K$  are treated as one input each. We are interested in the global influence of  $(\mathbf{X}, Z)$  where  $Z$  denotes  $Z^G$  or  $Z^K$  depending on the representation.

## 4.2 Numerical setting for sensitivity analysis

We consider a population of  $N = 2005$  individuals including five exposed individuals at the start of the epidemics  $t = 0$ , so that for all  $\theta \in \Theta$  the CTMC  $W^\theta$  has the initial state:

$$\xi_0 = (W_S(0) = 2000, W_E(0) = 5, W_A(0) = 0, W_I(0) = 0, \\ W_H(0) = 0, W_R(0) = 0, W_D(0) = 0).$$

For the quantity of interest  $Y_{ext}$ , the final time  $T$  of the study is set to  $T = +\infty$ . In practice, this means that trajectories of  $W^\theta$  are simulated until they reach absorbing states, i.e. extinction. Concerning  $Y_I$ ,  $T$  is set to  $T = 60$  days. This choice is motivated by the outputs of  $Y_I$  obtained with the nominal values of parameters and displayed in Figure S2. This figure shows that trajectories do not much vary after 60 days and about

two-thirds of them reach extinction right after  $T = 60$  days which is the 32th percentile of extinction times (see boxplot of Figure S1). Therefore,  $Y_I$  is studied on the interval  $[0, 60]$  that is discretized into 1000 equidistant time points for GSA purposes.

Regarding sensitivity index estimation, we draw  $n = 2000$  samples of each input with respect to uniform distributions over ranges specified in Table 2. The parameter space exploration is performed by using Latin Hypercube Sampling as the dimension of  $\Theta$  is quite large. We simulate  $W^\theta$  through Algorithm 2 for random time change representation and Algorithm 4 for Gillespie representation. In both cases, the intrinsic randomness variable is a vector of nine components. Nine integers are drawn independently and uniformly in  $\{1, \dots, 10^9\}$  to serve as seeds for the random number generator with the aim to stream random numbers.

In practice, simulations are carried out with *R*. We use the *R* package *DiceDesign* (Dupuy et al. [2015]) for Latin Hypercube Sampling. Sensitivity indices are estimated by using functions *soboljansen* and *sobol2007* of the *R* package *sensitivity* (Iooss et al. [2021]). The function *soboljansen* is used for total Sobol' index estimation while *sobol2007* is used for first-order Sobol' index estimation.

### 4.3 Sensitivity index estimation

GSA is performed for the two outputs. For  $Y_{ext}$ , first-order and total Sobol' indices of the nine inputs are estimated. For each representation, 50 replications of these indices are provided. Regarding  $Y_I$ , dynamical and aggregated Sobol' indices are estimated.

#### 4.3.1 Sensitivity analysis results for $Y_{ext}$

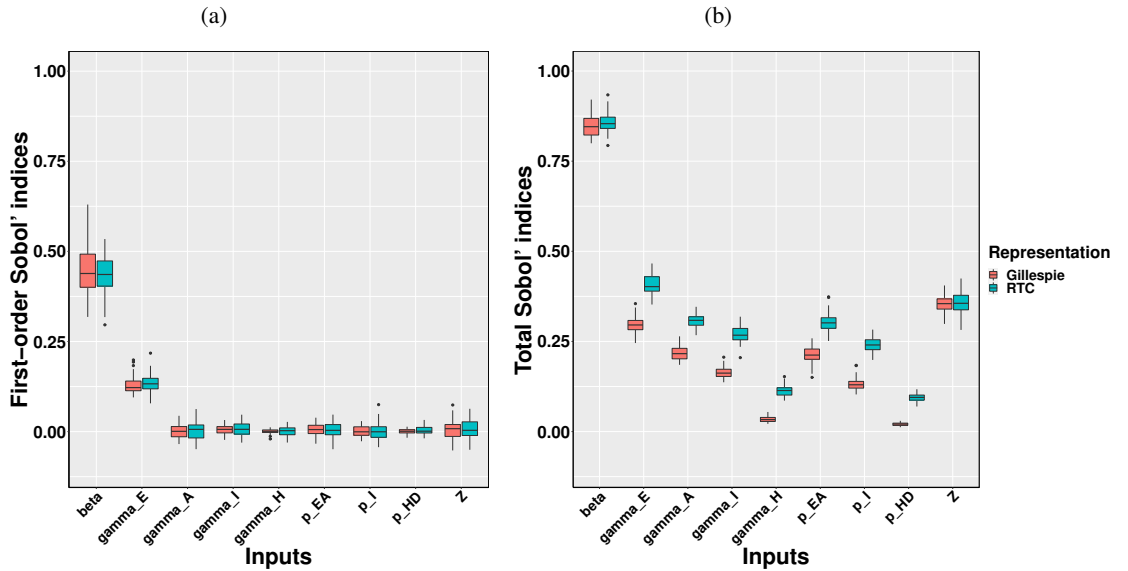


Figure 2: Boxplots of 50 replications of first-order Sobol' indices (a) and total Sobol' indices (b) for  $Y_{ext}$  with respect to Gillespie representation (red) and Random time change representation (RTC) (blue). Each index estimation is obtained with  $n = 2000$  samples of inputs.



Figure 2 suggests that, considering first-order indices, only two parameters influence the variability of  $Y_{\text{ext}}$ :  $\beta$  and  $\gamma_E$ . Also, there is no significant difference in the input sensitivity indices between the two model representations. Regarding total indices, results point out strong interactions between inputs. Thus, one can notice that almost all the inputs have non-zero total effects except  $p_{(H,D)}$  in the case of the Gillespie representation. Two main points can be emphasized: first, the total Sobol' index estimations of  $Z$  for the two representations are very close. Indeed, the total Sobol' index of  $Z$  is theoretically the same for the two representations (see Section 2.3). That is why estimations are equal up to sampling errors. Secondly, there are significant differences between the estimations for Gillespie and Random time change representations, and the three most influential inputs (more than 25% of variance each) are:  $\beta$ ,  $\gamma_E$  and  $Z$ . The intrinsic randomness interacts strongly with uncertain parameters. This explains the difference that appears between main and total effects.

The ranking of inputs with respect to total Sobol' indices for Gillespie representation yields  $\beta, Z, \gamma_E, \gamma_A, \gamma_I, p_{(E,A)}, p_I, \gamma_H, p_{(H,D)}$  while that of Random time change representation is  $\beta, \gamma_E, Z, \gamma_A, p_{(E,A)}, \gamma_I, p_I, \gamma_H, p_{(H,D)}$ . From one representation to another, rankings of inputs  $\gamma_E, Z, \gamma_I, p_{(E,A)}$  have switched. So, in practice, GSA conclusions relative to ranking depend on representation.

In order to confirm statistically the difference between indices, we carry out asymptotic statistical tests of comparison of means of two samples. For each input, the test consists in comparing the total Sobol' indices renormalized by the total variance (i.e. numerator of total Sobol' index formula defined in Subsection 2.1) obtained for the two representations. The samples derive from the total Sobol' index samples used for the boxplots in Figure 2.

The null hypothesis ( $H_0$ ) is rejected for inputs  $\gamma_E, \gamma_A, \gamma_I, \gamma_H, p_{EA}, p_I$  and  $p_{HD}$  with p-values less than  $2.2 \times 10^{-16}$  but not for inputs  $\beta$  and  $Z$  as their corresponding p-values 0.75 and 0.85 are greater than 5%. There is no surprise that the test for  $\beta$  is inconclusive since boxplots overlap (see Figure 2b).

Results provided in Subsection 4.3 reveal that differences can occur between GSA responses for the two representations. In particular, for total Sobol' indices of uncertain parameters, intrinsic randomness distributions are explicitly involved in formulas and consequently estimations for the two representations do not seem to be distributed around the same theoretical value. Even though these differences may legitimately exist, they can have significant repercussions in practice. For this model, we could indeed choose to adopt Gillespie representation as interactions with intrinsic noise are much less important for this representation. However, the most important message to keep in mind is that the intrinsic randomness varies from one representation to another, thus different representations yield different information. And in practice, this can impact GSA conclusions and even resulting decisions.

### 4.3.2 Sensitivity analysis results for $Y_I$

#### Dynamical Sobol' indices

As showed in Figure 3, both dynamical first-order and total Sobol' index estimation indicate that the most important input over time is  $\beta$ , followed by  $\gamma_E$ , and  $p_{(E,A)}$ , except at the very beginning of the epidemic where  $\gamma_E$  is the most important parameter. These conclusions are valid for both representations. First-order and total Sobol' indices of  $Z$  are the highest from the start until day 5, meaning that the intrinsic stochasticity influence exceeds that of uncertain parameters at the beginning of epidemics. This reflects

the fact that intrinsic stochasticity rules the dynamics especially at the beginning of outbreaks when the number of infectious is low.

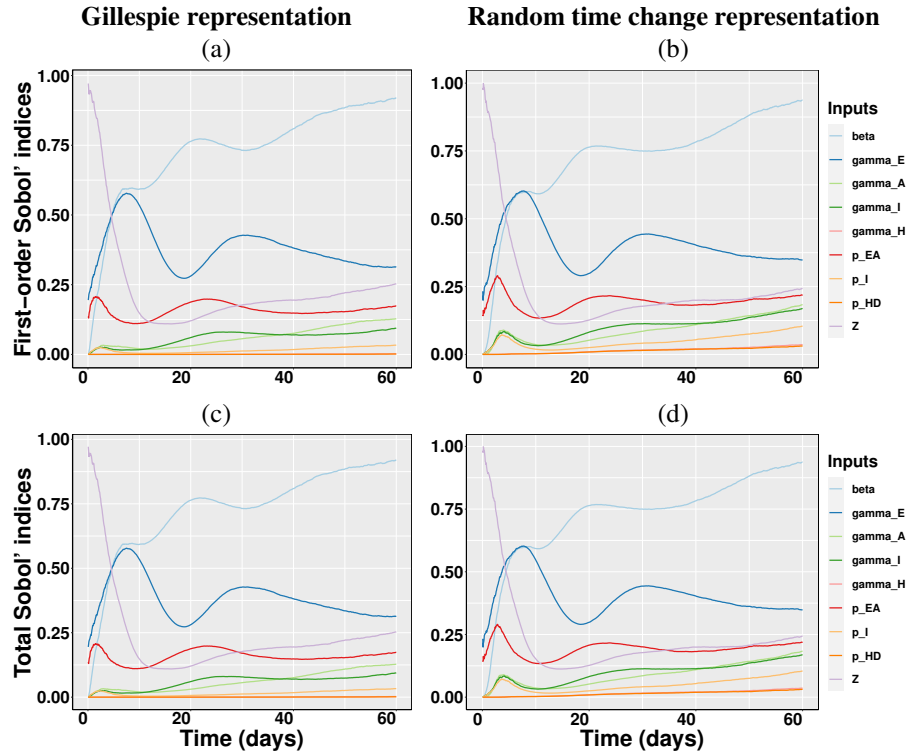


Figure 3: Dynamical first-order and total Sobol' indices for  $Y_I$  with respect to Gillespie representation (a) & (c) and Random time change representation (b) & (d). At each of the 1000 discretization time points in  $[0, 60]$ , first-order Sobol' indices are estimated with  $n = 2000$  samples of inputs.

### Aggregated Sobol' indices

The aggregated Sobol' indices enable to summarize over the whole time interval the impact of inputs. So, for both first-order and total indices, the transmission parameter  $\beta$  remains the most important input, followed by  $\gamma_E, p_{(E,A)}$  and  $Z$ . The impact of  $Z$  is again noticeable when considering total effects, due to its interactions with other inputs.

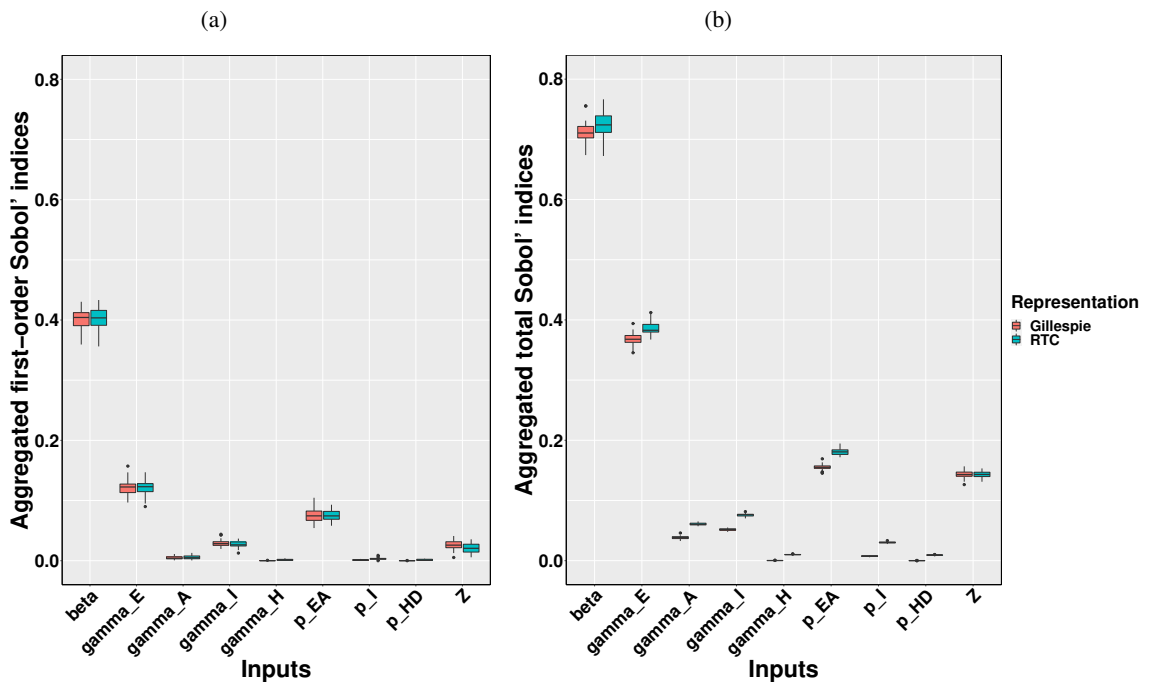


Figure 4: Aggregated first-order (a) and total (b) Sobol' indices for  $Y_T$  with respect to Gillespie representation (red) and Random time change representation (blue). Indices are obtained with the 1000 discretization time points in  $[0, 60]$  and  $n = 2000$  samples of inputs.

With the aim of validating the differences between total Sobol' indices for the two representations, statistical tests are carried out as in the case of the scalar output  $Y_{\text{ext}}$  by using samples that allowed to generate boxplots of Figure 4. The null hypothesis that states the equality between total Sobol' indices of input parameters for the two representations is rejected with p-values smaller than  $2.2 \times 10^{-16}$  except parameters  $\beta$  and  $\gamma_E$  for which the p-values are respectively  $2.24 \times 10^{-4}$  and  $2.6 \times 10^{-14}$ . The test for  $Z$  does not reject the null hypothesis with a p-value of 0.76. Overall, the conclusions of these tests perfectly match with the theoretical analysis detailed in Subsection 2.3.

## 5 Conclusions

In this work, we developed an approach of sensitivity analysis for stochastic compartmental models described by continuous-time Markov chains. This approach consists in separating intrinsic randomness from parameter uncertainty by building exact deterministic representations of the model outputs as functions of uncertain parameters and explicit intrinsic randomness. For this purpose, we rely on Gillespie SSA to propose a deterministic representation of continuous-time Markov chains. We present two versions of the new Gillespie SSA based representation: the direct method (see Proposition 2) and the first reaction method (see Proposition 3). The latter representation enables to highlight the impact of intrinsic noise of each type of transition or reaction channel of the model. Regarding the representation based on direct method, it has the advantage to be computationally faster and most commonly used in practice.

This approach is applied to a stochastic compartmental model of SARS-CoV-2 spread and is compared to an approach based on random time change analyzed in [Navarro Jimenez et al. \[2016\]](#). GSA is performed for two quantities of interest. The first QoI studied is the extinction time of epidemics. The Gillespie and random time change representations reveal that the transmission parameter ( $\beta$ ) is the most important input and that the intrinsic randomness ( $Z$ ) much interacts with uncertain parameters. The second QoI is functional and given by the dynamics of number of symptomatic infectious individuals. For both representations, sensitivity analysis highlights again the main role of transmission parameter ( $\beta$ ) and incubation mean duration parameter ( $\gamma_E$ ) and the slight influence of the intrinsic randomness ( $Z$ ), except at the very beginning of the epidemic. Estimating Sobol' indices of intrinsic randomness allowed to point out not only the different phases of the epidemic, regarding the influence of the intrinsic randomness, but also its global influence on the whole epidemic dynamic.

Overall, our approach allows to estimate sensitivity indices for the main effect of intrinsic randomness as well as for its interactions with uncertain parameters. This additional information is complementary to the one on the influence of uncertain parameters. We also highlighted the fact that GSA results depend on the chosen representation. In practice, this impacts conclusions resulting from GSA. Nevertheless, the different results between the representations provide interesting information; for instance, choosing a representation which has lower interaction between the intrinsic randomness and the uncertain parameters. This may allow to estimate uncertain parameter values with a better accuracy.

In future research, the approach proposed in this study could be extended to compartmental models based on non-Markovian stochastic processes based on Sellke construction ([Sellke \[1983\]](#)).

## Appendix

### A Some simulations of QoIs $Y_{ext}$ and $Y_I$

#### Case of $Y_{ext}$

For uncertain parameters set to nominal values (see [Table 2](#)), we simulate  $n = 2000$  outputs of  $Y_{ext}$ . The outcomes are presented under the form of boxplot in [Figure S1](#).

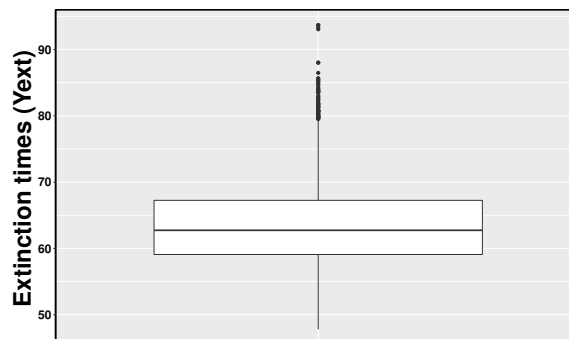


Figure S1: Boxplot of  $n = 2000$  outputs of  $Y_{ext}$

## Case of $Y_I$

Figure S2 shows 20 trajectories of the number of symptomatic infectious individuals corresponding to 20 outputs of the model for parameters set to nominal values (see Table 2).

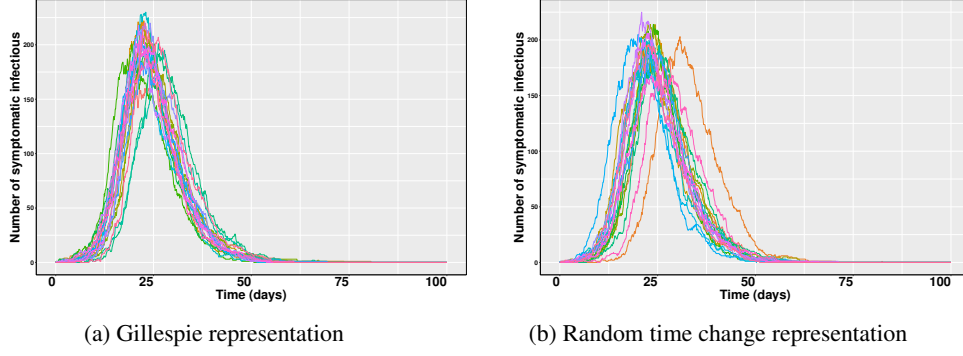


Figure S2: 20 trajectories of number of symptomatic infectious per representations

## B Proofs

### Proof of Proposition 1

The proof is provided for a case of stochastic process so that it can be adapted to real or multidimensional output stochastic models. Assume that for all  $\theta \in \Theta$ ,  $g(\theta)$  is a stochastic process  $\{g(\theta, t), t \geq 0\}$  over a discrete state space  $\mathcal{E}$ . Let  $A_1, \dots, A_d$  and  $B$  be Borel sets. Assume  $\mathbf{X}$  has a distribution  $\mu$ . Consider  $0 \leq t_1 \leq \dots \leq t_d$ . Note that  $\mathbb{P}(\mathbf{X} \in B, g(\mathbf{X}, t_1) \in A_1, \dots, g(\mathbf{X}, t_d) \in A_d) = \mathbb{E} \left[ \mathbf{1}_{\mathbf{X} \in B} \mathbb{E} \left[ \prod_{i=1}^d \mathbf{1}_{g(\mathbf{X}, t_i) \in A_i} \mid \mathbf{X} \right] \right]$ . Since  $\mathbb{E} \left[ \mathbf{1}_{\mathbf{X} \in B} \mathbb{E} \left[ \prod_{i=1}^d \mathbf{1}_{g(\mathbf{X}, t_i) \in A_i} \mid \mathbf{X} \right] \right] = \int_B \mathbb{E} \left[ \prod_{i=1}^d \mathbf{1}_{f(t_i, \theta, Z) \in A_i} \right] \mu(d\theta)$  then  $\mathbb{P}(\mathbf{X} \in B, g(\mathbf{X}, t_1) \in A_1, \dots, g(\mathbf{X}, t_d) \in A_d)$  is equal to  $\mathbb{P}(\mathbf{X} \in B, f(t_1, \mathbf{X}, Z) \in A_1, \dots, f(t_d, \mathbf{X}, Z) \in A_d)$ .

### Construction of Gillespie representation and Proof of Proposition 2

For all  $\theta \in \Theta$ ,  $W^\theta$  is a CTMC on the finite state space  $\mathcal{E}$  with generator  $\mathcal{Q}_\theta$ . The jump sequence  $\{T_n; n \geq 0\}$  of  $W^\theta$  is defined as:

$$\begin{cases} T_0 = 0 \\ T_{n+1} - T_n \sim \text{Exp} \left( \sum_{(\alpha, \beta) \in \mathbf{E}} g_{\alpha, \beta}(\theta, W^\theta(T_n)) \right) \\ \text{Conditionally to } \{W^\theta(T_n), n \geq 0\}, (T_{n+1} - T_n) \text{ are mutually independent} \end{cases}$$

The sequence  $\{W^\theta(T_n); n \geq 0\}$  is the embedded chain of  $W^\theta$  i.e. a discrete-time

Markov chain on  $\mathcal{E}$  with initial state  $\xi_0$  and transition probabilities given by:

$$\begin{aligned} \forall \xi, \xi' \in \mathcal{E}, p_{\xi, \xi'} &= \mathbb{P}(W^\theta(T_{n+1} = \xi' \mid W^\theta(T_n) = \xi)) \\ &= \begin{cases} \frac{g_{\alpha', \beta'}(\theta, \xi)}{\sum_{(\alpha, \beta) \in \mathbf{E}} g_{\alpha, \beta}(\theta, \xi)} & \text{if } \exists (\alpha', \beta') \in \mathbf{E} : \xi' - \xi = \mathbf{u}_{\alpha', \beta'} \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \quad (1)$$

Therefore,  $W^\theta$  can be rewritten under the form:

$$W^\theta(t) = \sum_{n=0}^{+\infty} W^\theta(T_n) \cdot \mathbb{1}_{[T_n, T_{n+1}[}(t).$$

Consider a discrete-time Markov chain  $D^\theta = \{D_n^\theta; n \geq 0\}$  with transition probabilities  $(p_{\xi, \xi'})_{(\xi, \xi') \in \mathcal{E} \times \mathcal{E}}$  defined in Eq. (1) and initial state  $\xi_0$ . If the sequence  $\{\tau_n; n \geq 0\}$  satisfies:

$$\begin{cases} \tau_0 = 0 \\ \tau_{n+1} - \tau_n \sim \text{Exp}\left(\sum_{(\alpha, \beta) \in \mathbf{E}} g_{\alpha, \beta}(\theta, D_n^\theta)\right) \\ \text{Conditionally to } D^\theta, (\tau_{n+1} - \tau_n) \text{ are mutually independent} \end{cases} \quad (2)$$

then the stochastic process  $H^\theta = \{\sum_{n=0}^{+\infty} D_n^\theta \cdot \mathbb{1}_{[\tau_n, \tau_{n+1}[}(t); t \geq 0\}$  defines a continuous-time Markov chain with generator  $\mathcal{Q}_\theta$ , state space  $\mathcal{E}$  and initial state  $\xi_0$ . Since  $\mathcal{E}$  is finite then:

$$\forall \theta \in \Theta, \quad W^\theta \stackrel{f.d.d.}{=} H^\theta. \quad (3)$$

Equation (3) implies that the two processes are stochastically equivalent. The discrete chain  $D^\theta$  and sequence  $\{\tau_n; n \geq 0\}$  can be arbitrarily chosen. Let us construct an example of  $D^\theta$  and sequence  $\{\tau_n; n \geq 0\}$  using two independent sequences  $\mathbf{U}^1 = (U_1^1, U_2^1, \dots)$  and  $\mathbf{U}^2 = (U_1^2, U_2^2, \dots)$  of i.i.d. standard uniform variables.

Note that  $\mathbf{E}$  is finite as  $\mathbf{V}$  is finite. Recall that  $n_E$  the cardinal of  $\mathbf{E}$ . So, we order the elements of  $\mathbf{E}$  with respect to an arbitrarily order:  $\mathbf{E} = \{(\alpha^{(1)}, \beta^{(1)}), \dots, (\alpha^{(n_E)}, \beta^{(n_E)})\}$ . Let us introduce the function:

$$\begin{aligned} \phi^\theta : \mathcal{E} \times [0, 1] &\longrightarrow \mathcal{E} \\ (\xi, v) &\longmapsto \xi + \sum_{k=1}^{n_E} \mathbf{u}_{\alpha^{(k)}, \beta^{(k)}} \times \mathbb{1}_{v \in \left[ \sum_{j=1}^{k-1} p_{\xi, \xi + \mathbf{u}_{\alpha^{(j)}, \beta^{(j)}}}, \sum_{j=1}^k p_{\xi, \xi + \mathbf{u}_{\alpha^{(j)}, \beta^{(j)}}} \right]} \end{aligned} \quad (4)$$

The discrete chain built by the recursion:

$$\begin{cases} D_0 = \xi_0 \\ D_n^\theta = \phi^\theta(D_{n-1}^\theta, U_n^1), n \geq 1 \end{cases} \quad (5)$$

is a Markov chain with transition probabilities  $(p_{\xi, \xi'})_{(\xi, \xi') \in \mathcal{E} \times \mathcal{E}}$ . Furthermore, the sequence:

$$\begin{cases} \tau_0 = 0 \\ \tau_{n+1} - \tau_n = -\frac{\log(U_n^2)}{\sum_{(\alpha, \beta) \in \mathbf{E}} g_{\alpha, \beta}(\theta, D_n^\theta)} \end{cases} \quad (6)$$

satisfies conditions of Eq. (2.) Therefore the discrete chain in Eq. (5) and the sequence in Eq. (6) define a CTMC stochastically equivalent to  $W^\theta$  in the sense of finite dimensional distributions.

Let define the function:

$$\begin{aligned} \Phi^\theta : \mathcal{L}([0, 1]) &\longrightarrow \mathcal{E}^\mathbb{N} \\ u = (u_1, u_2, \dots) &\longmapsto \Phi^\theta(u) = (\xi_0, \phi^\theta(\xi_0, u_1), \phi^\theta(\phi^\theta(\xi_0, u_1), u_2), \dots) \end{aligned}$$

By construction  $\Phi^\theta(\mathbf{U}^1) = \{\Phi_n^\theta(\mathbf{U}^1) = D_n^\theta, n \geq 0\}$ . If  $\lambda_n = \sum_{(\alpha, \beta) \in \mathbf{E}} g_{\alpha, \beta}(\theta, \Phi_n^\theta(\mathbf{U}^1))$  then the following function:

$$\begin{aligned} f_G : \mathbb{R}_+ \times \Theta \times \mathcal{L}([0, 1])^2 &\longrightarrow \mathcal{E} \\ (t, \theta, z_1, z_2) &\longmapsto f_G(t, \theta, z_1, z_2) = \sum_{n=0}^{\infty} \Phi_n^\theta(z_1) \mathbb{1}_{t \in [\sum_{j=1}^n \frac{-\log(z_{2,j})}{\lambda_{j-1}}, \sum_{j=1}^{n+1} \frac{-\log(z_{2,j})}{\lambda_{j-1}})} \end{aligned}$$

is such that  $f_G(\cdot, \theta, Z^G)$  is stochastically equivalent to  $W^\theta$  with  $Z^G = (\mathbf{U}^1, \mathbf{U}^2)$ .

Moreover, since  $\mathbf{X}$  and  $Z^G$  are independent and  $f_G(\cdot, \theta, Z^G) \sim W^\theta$  then the equivalence  $(\mathbf{X}, \mathbf{W}) \sim (\mathbf{X}, f_G(\cdot, \mathbf{X}, Z^G))$  is proved in the same way as in Proposition 1.

### Second construction of Gillespie representation and Proof of Proposition 3

We consider the vector of sequences of i.i.d. standard random variables:  $(\mathbf{U}_{\alpha, \beta}, (\alpha, \beta) \in \mathbf{E})$  where  $\mathbf{U}_{\alpha, \beta} = (U_{\alpha, \beta}^1, U_{\alpha, \beta}^2, \dots)$ . Given the following recursive system:

$$\begin{cases} D_0^\theta = \xi_0 \\ s_{\alpha, \beta}^n = -\frac{\log(U_{\alpha, \beta}^n)}{g_{\alpha, \beta}(\theta, D_{n-1}^\theta)} \quad \forall (\alpha, \beta) \in \mathbf{E} \\ D_n^\theta = D_{n-1}^\theta + \mathbf{u}_{\alpha, \beta}^- \mathbb{1}_{s_{\alpha, \beta}^n = \min\{s_{\alpha, \beta}^n; (\alpha, \beta) \in \mathbf{E}\}} \end{cases} \quad (7)$$

$(\mathbf{U}_{\alpha, \beta}, (\alpha, \beta) \in \mathbf{E})$  is a vector of sequences of i.i.d. standard random variables where  $\mathbf{U}_{\alpha, \beta} = (U_{\alpha, \beta}^1, U_{\alpha, \beta}^2, \dots)$ . We have the following system:

$$\begin{cases} D_0^\theta = \xi_0 \\ s_{\alpha, \beta}^n = -\frac{\log(U_{\alpha, \beta}^n)}{g_{\alpha, \beta}(\theta, D_{n-1}^\theta)} \quad \forall (\alpha, \beta) \in \mathbf{E} \\ D_n^\theta = D_{n-1}^\theta + \mathbf{u}_{\alpha, \beta}^- \mathbb{1}_{s_{\alpha, \beta}^n = \min\{s_{\alpha, \beta}^n; (\alpha, \beta) \in \mathbf{E}\}} \end{cases}$$

1. Set  $U_n = \{U_{\alpha, \beta}^n; (\alpha, \beta) \in \mathbf{E}\}$ .  $\{U_n; n \geq 1\}$  is a sequence of i.i.d. random vectors with distribution  $\otimes_{i=1}^{n_E} \mathcal{U}([0, 1])$ . Given the function:

$$\begin{aligned} \varphi^\theta : \mathcal{E} \times [0, 1]^{n_E} &\longrightarrow \mathcal{E} \\ (\xi, u = u_{\alpha, \beta}^n; (\alpha, \beta) \in \mathbf{E}) &\longmapsto \xi + \mathbf{u}_{\alpha, \beta}^- \mathbb{1}_{s_{\alpha, \beta}^n = \min\{-\frac{\log(u_{\alpha, \beta}^n)}{g_{\alpha, \beta}(\theta, \xi)}; (\alpha, \beta) \in \mathbf{E}\}} \end{aligned}$$

it yields  $D_n = \varphi^\theta(D_{n-1}, U_n)$ . So,  $\{D_n^\theta, n \geq 0\}$  defines a Markov chain with initial state  $\xi_0$ . Let us find the transition probabilities of this chain. Let  $\xi'$  and  $\xi$  be two states such that  $\xi' \neq \xi$ . We shall discuss two cases.

- Case 1:  $\xi' \neq \xi + \mathbf{u}_{\alpha,\beta}$  for all  $(\alpha, \beta) \in \mathbf{E}$ . In this case, we have:

$$\begin{aligned} \mathbb{P}(D_{n+1}^\theta = \xi' \mid D_n^\theta = \xi) &= \mathbb{P}\left(\xi' = \xi + \mathbf{u}_{\bar{\alpha},\bar{\beta}} \mathbb{1}_{s_{\bar{\alpha},\bar{\beta}}^n = \min\{s_{\alpha,\beta}^n; (\alpha,\beta) \in \mathbf{E}\}} \mid D_n^\theta = \xi\right) \\ &= \mathbb{P}\left(\xi' = \xi + \mathbf{u}_{\bar{\alpha},\bar{\beta}} \mathbb{1}_{s_{\bar{\alpha},\bar{\beta}}^n = \min\{s_{\alpha,\beta}^n; (\alpha,\beta) \in \mathbf{E}\}}\right) \\ &= 0 \end{aligned}$$

- Case 2:  $\exists (\hat{\alpha}, \hat{\beta}) \in \mathbf{E}$  such that  $\xi' = \xi + \mathbf{u}_{\hat{\alpha},\hat{\beta}}$ . In this case, we have:

$$\begin{aligned} \mathbb{P}(D_n^\theta = \xi' \mid D_{n-1}^\theta = \xi) &= \mathbb{P}\left(\xi' = \xi + \mathbf{u}_{\bar{\alpha},\bar{\beta}} \mathbb{1}_{s_{\bar{\alpha},\bar{\beta}}^n = \min\{s_{\alpha,\beta}^n; (\alpha,\beta) \in \mathbf{E}\}} \mid D_{n-1}^\theta = \xi\right) \\ &= \mathbb{P}\left(\mathbf{u}_{\hat{\alpha},\hat{\beta}} = \mathbf{u}_{\bar{\alpha},\bar{\beta}} \mathbb{1}_{s_{\bar{\alpha},\bar{\beta}}^n = \min\{s_{\alpha,\beta}^n; (\alpha,\beta) \in \mathbf{E}\}} \mid D_{n-1}^\theta = \xi\right) \\ &= \mathbb{P}\left(s_{\hat{\alpha},\hat{\beta}}^n = \min\{s_{\alpha,\beta}^n; (\alpha, \beta) \in \mathbf{E}\} \mid D_{n-1}^\theta = \xi\right) \\ &= \mathbb{P}\left(s_{\hat{\alpha},\hat{\beta}}^n \leq s_{\alpha,\beta}^n; (\alpha, \beta) \in \mathbf{E} \setminus (\hat{\alpha}, \hat{\beta}) \mid D_{n-1}^\theta = \xi\right) \\ &= \mathbb{E}\left[\exp\left(-s_{\hat{\alpha},\hat{\beta}}^n \times \sum_{(\alpha,\beta) \in \mathbf{E} \setminus (\hat{\alpha},\hat{\beta})} g_{\alpha,\beta}(\theta, \xi)\right) \mid D_{n-1}^\theta = \xi\right] \\ &= \frac{g_{\hat{\alpha},\hat{\beta}}(\theta, \xi)}{\sum_{(\alpha,\beta) \in \mathbf{E}} g_{\alpha,\beta}(\theta, \xi)} \end{aligned}$$

It turns out that  $\{D_n^\theta, n \geq 0\}$  has transition matrix  $(p_{\xi,\xi'}_{(\xi,\xi') \in \mathcal{E} \times \mathcal{E}})$ .

2. It is straightforward since  $\{U_{\alpha,\beta}^n; (\alpha, \beta) \in \mathbf{E}\}$  are i.i.d. and  $s_{\alpha,\beta}^n$  are exponential variables.
3. The construction of  $f_G$  is similar to the one of the proof of Proposition 2 by using  $Z^G = (\mathbf{U}_{\alpha,\beta}, (\alpha, \beta) \in \mathbf{E})$  and  $\phi^\theta = \varphi^\theta$ .

## References

- D. F. Anderson. A modified next reaction method for simulating chemical systems with time dependent propensities and delays. *The Journal of Chemical Physics*, 127(21):214107, 2007. doi: 10.1063/1.2799998. URL <https://doi.org/10.1063/1.2799998>.
- B. Cazelles, C. Champagne, B. Nguyen-Van-Yen, C. Comiskey, E. Vergu, and B. Roche. A mechanistic and data-driven reconstruction of the time-varying reproduction number: Application to the covid-19 epidemic. *PLOS Computational Biology*, 17(7):1–20, 07 2021. doi: 10.1371/journal.pcbi.1009211. URL <https://doi.org/10.1371/journal.pcbi.1009211>.
- A. Courcoul, H. Monod, M. Nielen, D. Klinkenberg, L. Hogerwerf, F. Beaudau, and E. Vergu. Modelling the effect of heterogeneity of shedding on the within herd coxiella burnetii spread and identification of key parameters by sensitivity analysis. *Journal of Theoretical Biology*, 284(1):130–141, 2011. ISSN 0022-5193. doi: <https://doi.org/10.1016/j.jtbi.2011.06.017>. URL <https://www.sciencedirect.com/science/article/pii/S002251931100316X>.



- L. Cristancho Fajardo, P. Ezanno, and E. Vergu. Accounting for farmers' control decisions in a model of pathogen spread through animal trade. *Scientific Reports*, 11 (1):9581, May 2021. ISSN 2045-2322. doi: 10.1038/s41598-021-88471-6. URL <https://doi.org/10.1038/s41598-021-88471-6>.
- S. da Veiga. Kernel-based anova decomposition and shapley effects – application to global sensitivity analysis, 2021. arXiv:2101.05487.
- D. Dupuy, C. Helbert, and J. Franco. DiceDesign and DiceEval: Two R packages for design and analysis of computer experiments. *Journal of Statistical Software*, 65 (11):1–38, 2015. URL <https://www.jstatsoft.org/v65/i11/>.
- S. N. Ethier and T. 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.
- P. Étoré, C. Prieur, D. K. Pham, and L. Li. Global Sensitivity Analysis for Models Described by Stochastic Differential Equations. *Methodology and Computing in Applied Probability*, 22(2):803–831, June 2020. ISSN 1387-5841, 1573-7713. doi: 10.1007/s11009-019-09732-6. URL <http://link.springer.com/10.1007/s11009-019-09732-6>.
- J.-C. Fort, T. Klein, and A. Lagnoux. Global sensitivity analysis and wasserstein spaces. *SIAM/ASA Journal on Uncertainty Quantification*, 9(2):880–921, 2021. doi: 10.1137/20M1354957. URL <https://doi.org/10.1137/20M1354957>.
- F. Gamboa, A. Janon, T. Klein, and A. Lagnoux. Sensitivity analysis for multidimensional and functional outputs. *Electronic Journal of Statistics*, 8(1):575 – 603, 2014. doi: 10.1214/14-EJS895. URL <https://doi.org/10.1214/14-EJS895>.
- D. 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](https://doi.org/10.1016/0021-9991(76)90041-3). URL <https://www.sciencedirect.com/science/article/pii/0021999176900413>.
- J. L. Hart, A. Alexanderian, and P. A. Gremaud. Efficient computation of sobol' indices for stochastic models. *SIAM Journal on Scientific Computing*, 39(4):A1514–A1530, 2017. doi: 10.1137/16M106193X. URL <https://doi.org/10.1137/16M106193X>.
- T. Homma and A. Saltelli. Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering and System Safety*, 52(1):1–17, 1996. ISSN 0951-8320. doi: [https://doi.org/10.1016/0951-8320\(96\)00002-6](https://doi.org/10.1016/0951-8320(96)00002-6). URL <https://www.sciencedirect.com/science/article/pii/0951832096000026>.
- B. Iooss, S. D. Veiga, A. Janon, G. Pujol, with contributions from Baptiste Broto, K. Boumhaout, T. Delage, R. E. Amri, J. Fruth, L. Gilquin, J. Guillaume, M. I. Idrissi, L. Le Gratiet, P. Lemaitre, A. Marrel, A. Meynaoui, B. L. Nelson, F. Monari, R. Oomen, O. Rakovec, B. Ramos, O. Roustant, E. Song, J. Staum, R. Sueur, T. Touati, and F. Weber. *sensitivity: Global Sensitivity Analysis of Model Outputs*, 2021. URL <https://CRAN.R-project.org/package=sensitivity>. R package version 1.24.0.

- M. N. Jimenez, O. P. Le Maître, and O. M. Knio. Nonintrusive Polynomial Chaos Expansions for Sensitivity Analysis in Stochastic Differential Equations. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):378–402, Jan. 2017. ISSN 2166-2525. doi: 10.1137/16M1061989. URL <https://epubs.siam.org/doi/10.1137/16M1061989>.
- E. S. Knock, L. K. Whittles, J. A. Lees, P. N. Perez-Guzman, R. Verity, R. G. FitzJohn, K. A. Gaythorpe, N. Imai, W. Hinsley, L. C. Okell, A. Rosello, N. Kantas, C. E. Walters, S. Bhatia, O. J. Watson, C. Whittaker, L. Cattarino, A. Boonyasiri, B. A. Djaafara, K. Fraser, H. Fu, H. Wang, X. Xi, C. A. Donnelly, E. Jauneikaite, D. J. Laydon, P. J. White, A. C. Ghani, N. M. Ferguson, A. Cori, and M. Baguelin. The 2020 sars-cov-2 epidemic in england: key epidemiological drivers and impact of interventions. *medRxiv*, 2021. doi: 10.1101/2021.01.11.21249564. URL <https://www.medrxiv.org/content/early/2021/01/13/2021.01.11.21249564>.
- T. G. Kurtz. Representation and approximation of counting processes. In W. H. Fleming and L. G. Gorostiza, editors, *Advances in Filtering and Optimal Stochastic Control*, pages 177–191, Berlin, Heidelberg, 1982. Springer Berlin Heidelberg. ISBN 978-3-540-39517-1.
- M. Lamboni, M. Hervé, and M. 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>. URL <http://www.sciencedirect.com/science/article/pii/S0951832010002504>.
- O. Le Maître and O. Knio. Pc analysis of stochastic differential equations driven by wiener noise. *Reliability Engineering & System Safety*, 135:107–124, 2015. ISSN 0951-8320. doi: <https://doi.org/10.1016/j.ress.2014.11.002>. URL <https://www.sciencedirect.com/science/article/pii/S0951832014002749>.
- 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. URL <https://doi.org/10.1063/1.4922922>.
- A. Marrel, B. Iooss, S. Da Veiga, and M. Ribatet. Global sensitivity analysis of stochastic computer models with joint metamodels. *Statistics and Computing*, 22(3):833–847, May 2012. ISSN 1573-1375. doi: 10.1007/s11222-011-9274-8. URL <https://doi.org/10.1007/s11222-011-9274-8>.
- G. Mazo. A trade-off between explorations and repetitions for estimators of two global sensitivity indices in stochastic models induced by probability measures. *SIAM/ASA Journal on Uncertainty Quantification*, 9(4):1673–1713, 2021. doi: 10.1137/19M1272706. URL <https://doi.org/10.1137/19M1272706>.
- M. Navarro Jimenez, O. P. Le Maître, and O. M. Knio. Global sensitivity analysis in stochastic simulators of uncertain reaction networks. *The Journal of Chemical Physics*, 145(24):244106, 2016. doi: 10.1063/1.4971797. URL <https://doi.org/10.1063/1.4971797>.
- Q. Richard, S. Alizon, M. Choisy, M. T. Sofonea, and R. Djidjou-Demasse. Age-structured non-pharmaceutical interventions for optimal control of covid-19 epidemic. *PLOS Computational Biology*,

- 17(3):1–25, 03 2021. doi: 10.1371/journal.pcbi.1008776. URL <https://doi.org/10.1371/journal.pcbi.1008776>.
- L. Rimbaud, C. Bruchou, S. Dallot, D. R. J. Pleydell, E. Jacquot, S. Soubeyrand, and G. Thébaud. Using sensitivity analysis to identify key factors for the propagation of a plant epidemic. *Royal Society Open Science*, 5(1):171435, 2018. doi: 10.1098/rsos.171435. URL <https://royalsocietypublishing.org/doi/abs/10.1098/rsos.171435>.
- T. 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. URL <https://ci.nii.ac.jp/naid/10027137978/en/>.
- X. Zhu and B. Sudret. Global sensitivity analysis for stochastic simulators based on generalized lambda surrogate models. *Reliability Engineering & System Safety*, 214:107815, 2021. ISSN 0951-8320. doi: <https://doi.org/10.1016/j.ress.2021.107815>. URL <https://www.sciencedirect.com/science/article/pii/S0951832021003379>.