



**HAL**  
open science

# A Tutorial on Sobol' Global Sensitivity Analysis Applied to Biological Models

Michel Tosin, Adriano M A Côrtes, Americo Cunha Jr

## ► To cite this version:

Michel Tosin, Adriano M A Côrtes, Americo Cunha Jr. A Tutorial on Sobol' Global Sensitivity Analysis Applied to Biological Models. da Silva F.A.B., Carels N., Trindade dos Santos M., Lopes F.J.P. Networks in Systems Biology: Applications for Disease Modeling, Springer, Cham, 2020, Mechanisms and Machine Science, 10.1007/978-3-030-51862-2\_6 . hal-02967410

**HAL Id: hal-02967410**

**<https://hal.science/hal-02967410>**

Submitted on 14 Oct 2020

**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.

Copyright

# Chapter 1

## A Tutorial on Sobol' Global Sensitivity Analysis Applied to Biological Models

Michel Tosin, Adriano M. A. Côrtes, and Americo Cunha Jr

**Abstract** Nowadays, in addition to traditional qualitative methods, quantitative techniques are also a standard tool to describe biological systems behavior. An example is the broad class of mathematical models, based on differential equations, used in ecology, biochemical kinetics, epidemiology, gene regulatory networks, etc. Independent of their simplicity or complexity, all these models have in common (generally unknown a priori) parameters that need to be identified from observations (data) of the real system, usually available on the literature, obtained by specific assays or surveyed by public health offices. Before using this data to calibrate the models, a good practice is to judge the most influential parameters. That can be done with aid of the Sobol' indices, a variance-based statistical technique for global sensitivity analysis, which measures the individual importance of each parameter, as well as their joint-effect, on the model output (a.k.a. quantity of interest). These variance-based indexes may be computed using Monte Carlo simulation but, depending on the model, this task can be very costly. An alternative approach for this scenario is the use of surrogate models to speed-up the calculations. Using simple biological models, from different areas, we develop a tutorial that illustrates how practitioners can use Sobol' indices to quantify, in a probabilistic manner, the relevance of the parameters of their models. This tutorial describes a very robust framework to compute Sobol' indices employing a polynomial chaos surrogate model constructed with the UQLab package.

---

Michel Tosin and Americo Cunha Jr  
Rio de Janeiro State University – UERJ, Nucleus of Modeling and Experimentation with Computers  
– NUMERICO, Rua São Francisco Xavier, 524, Rio de Janeiro, RJ - 20550-900, Brazil  
e-mail: [michel.tosin@uerj.br](mailto:michel.tosin@uerj.br); [americo@ime.uerj.br](mailto:americo@ime.uerj.br)

Adriano M. A. Côrtes  
Federal University of Rio de Janeiro – UFRJ, Multidisciplinary Computer Research Center –  
NUMPEX-COMP, Rodovia Washington Luiz, Km 104.5, Duque de Caxias, RJ - 25.265-970,  
Brazil  
e-mail: [adriano@nacad.ufrj.br](mailto:adriano@nacad.ufrj.br)

**Key words:** Mathematical biology; Global Sensitivity Analysis; Sobol' Indices; Surrogate Model; Polynomial Chaos Expansion; UQLab

## 1.1 Introduction

The progress that computer models have made over the past few decades in various areas of biology is impressive, with an increasing demand for the use of these tools in several scenarios that require quantitative predictions, such as infectious diseases [1, 2], change in species dynamics due to climate change [3, 4], protein signaling in cells [5, 6], etc. For instance, the recent pandemic of the Coronavirus SARS-CoV-2, responsible for the disease COVID-19 [7, 8]. The rapid spread of the disease is problematic because of the risk of death, and also for most of the health systems are not prepared to receive so many sick people in hospitals [9]. This kind of emergency scenario illustrates the importance of cooperation between researchers from different areas to deal with these situations by creating some understanding about the disease and propose strategies to slow down its reach, as fast as possible. The use of mathematical models can be very useful in situations like this, as long as it has the necessary ingredients to describe the basic aspects of the biological system of interest, and it is well-calibrated with real data [10].

The utility of a model must precede a process of certification to guarantee its match with reality [11]. During this process, it is useful to understand how certain variations in a model input parameter can affect its outcome (response) [12]. This knowledge allows one to detect which phenomena are more important in the real system [13, 14]. For instance, in the Coronavirus example, the effect of the response time to the outbreak by local governments and the various mitigation strategies has been evaluated [9].

In this sense, the mathematical model parameters are the “fundamental blocks” of the predictive tool and understand how each of these “pieces” affects the outcomes related to the biological system of interest is a key point for proper use of this quantitative arsenal. Global sensitivity analysis can be very useful to clarify this understanding, once it can provide information about the dependence of the model outcome with respect to each one of its input parameters [15]. Following that idea, this chapter aims to present a tutorial that illustrates the process of global sensitivity analysis in biological systems via Sobol' indices [16].

In Section 1.2 we present the setting for the mathematical model representation. Sobol' indices method is described in the Section 1.3, where the sensitivity analysis ideas and goals are also characterized. After that, Section 1.4 brings a brief overview of surrogate models as a computationally efficient strategy to approximate the Sobol' indices. With the theory detailed, Section 1.5 is responsible for presenting the Sobol framework and developing three biological examples: the predator-prey model, an NF- $\kappa$ B signaling pathway model and the SIR epidemiological model. All of them reproducible by the reader applying the codes available in a public repository.

## 1.2 Mathematical modeling

Modeling is the process of creating a (simplified) representation of the real process [17]. The objective is not being completely faithful to reality, but just being able to present its main characteristics or facilitate the identification and understanding of its mechanisms and processes [10]. To that goal, we can make several types of models for a single reality. With a good model, you can analyze many things about your problem to obtain more knowledge about it. Between the modeling objectives, we can highlight the understanding of the mechanisms that govern the phenomena of interest, prediction of the future or of some state that is currently unknown, and control by constraining a system to produce a desirable condition [18]. Because of this, mathematical modeling is often required to deal with biological processes.

The process of modeling involves some paradigms and steps but, more objectively, any model is composed by three elements: (i) the input, that comes from the previous useful information about the system state; (ii) the output, which is the information about the real problem; (iii) the model, which is the system representation that maps input to the output [19]. We can make a parallel with a manufactured production defining the raw material as the input, the processed product as the output and the machine responsible for transform one in another as the model. Intuitively, with a more sophisticated machine, we can create a more elaborate object as well as with a more detailed model we can obtain a better representation of the reality. However, including further details is necessary for a more complex representation. Greater complexity implies greater difficulty in analyzing all aspects of interest. Thereby, the hypotheses must balance the most important elements of the phenomena and a reasonable form of representing the relationship of these elements with what is interesting to quantify or evaluate. The model itself is this relation [20]. For example, the evolution of some animal species is related to the food resources available but also with the topographic characteristics of the environment. Assuming the species has lived in an environment for a long time, we can neglect then it and only take in count the first one. Sure, the effect of this approximation must be tested to guarantee your model is sufficiently compatible with reality. Additionally, the number of individuals of the species will increases (or decreases) according to some units of time. The change in a month is probably different than in a year. This time scale relation may be relevant or not, depending on the process treated.

Technically we represent the important input components of a real system in the model by parameters. In mathematical words, these parameters are quantities whose values are dependent on specific aspects of the phenomena and that controls how the model entails the desired predictions [10]. This prediction is called quantity (or quantities if are more than one) of interest (QoI) . So, we can think of the mathematical model as a “machine” that uses the parameters as ingredients to obtain the QoIs . Formally, the model is defined by a mathematical operator  $\mathcal{M}$ , so that the QoIs are given by

$$\mathbf{y} = \mathcal{M}(\mathbf{x}), \quad (1.1)$$

where the vector  $\mathbf{x} \in \mathbb{R}^m$  contains all the model parameters and the  $n$  QoIs are reunited in a vector  $\mathbf{y} \in \mathbb{R}^n$ . The structure of a general computational model is illustrated in Fig. 1.1. If the model is time dependent, one can lump the QoI at the instants of interest into a vector. To simplify the notation, the time dependence will be omitted in the theoretical-based discussions.

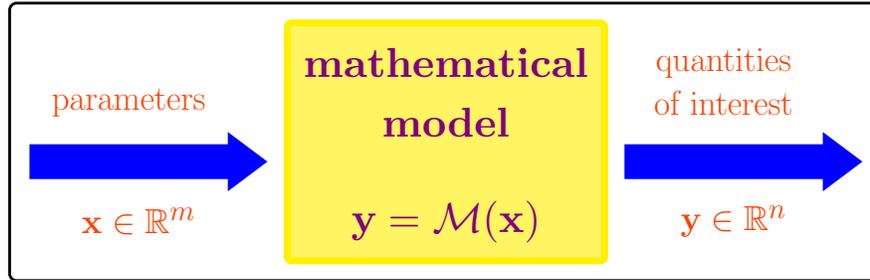


Fig. 1.1 Schematic representation of a general computational model.

### 1.3 Sensitivity Analysis

Sensitivity Analysis (SA) is a process in which the contribution of each input parameter of a mathematical model, to its response, is identified [21]. In particular, for nonlinear systems, not only the parameters individually affect the QoIs, but also the interactions between them. In this case, the joint-effect of parameter variation must also be quantified. This feature makes difficult the process of SA for high dimensional systems because several orders of interactions must be computed, and their influences are not trivial in general. Another important issue in the modeling process is to take into account the underlying uncertainties. The lack of knowledge about the real system and the natural variability of some parameters create some difficulties in any attempt to understand a phenomenon, that can be compensated by the use of a stochastic model [14, 19]. This is the domain of investigation of the Uncertainty Quantification (UQ) theory.

UQ and SA areas gained greater notoriety in the last decades, where it was recognized that is extremely important not only to deal with multiple sources of uncertainties in a mathematical model but, in some contexts, also apply strategies to reduce their effects, especially if their presence can be catastrophic for the system [22]. An alternative approach is also possible, where one tries to take advantage of the uncertainties to improve the performance of a given system. Although UQ and SA literature are already considerable, it is common to find texts that confuse these two concepts [13]. The first one deals with the quantitative evaluation of how the uncertainties in the input variables are transported to the model response, which is done through the calculation of the so-called uncertainty propagation process. On

the other side, the second concept is related to the quantitative evaluation of how uncertainty in the input parameters contribute (individually or jointly) to change the model response. In UQ it is mandatory to prescribe a characterization of the uncertainties so that after the calculation of their propagation, one has a general overview of the uncertainty in the model's response. In the SA perspective, the idea is to discover how a certain parameter changes the system response when changed [23].

The SA methods can be distinguished in two large groups: The Local SA (LSA) methods and the Global SA (GSA) methods [24]. From a general perspective, the difference between them is how the parameter space is explored in each case. Local methods are normally based in partial derivatives or gradients but are (totally) dependent on the point of the parameter space for which the model is evaluated. In this way, the LSA results do not reflect the general dependency of model outcome concerning the selected input parameter. Moreover, for nonlinear models, the interactions of the parameters are very important for the response and local analysis can not capture those effects. Differently, GSA methods explore screening or variance decomposition to cover the local analysis limitations [23]. In this chapter the focus of our interest in GSA, as a form to discover the most important parameters, that is, those that contribute the most to the model response (factor prioritization) and those which contribute very little and can potentially be fixed (factor fixing) [13].

### 1.3.1 Sobol' indices

There are several methods to perform SA present in scientific texts [13, 14]. But as said in the previous section, here we have preference for global methods. For the global sensitivity analysis framework to be robust and general, it makes sense to select a method that is simple to implement and use. The Sobol' indices [16] is a variance-based method very popular in recent literature [12] after the dissemination of surrogate model ideas, and very soon will be clear why.

In this framework, the system is analyzed from a probabilistic perspective, that considers the model input as a random vector  $\mathbf{X}$ , characterized by a joint probability density function (PDF)  $f_{\mathbf{X}}$  with support  $I_{\mathbf{X}}$ . The stochastic version of the model is represented as

$$\mathbf{Y} = \mathcal{M}(\mathbf{X}), \quad (1.2)$$

which has joint PDF  $f_{\mathbf{Y}}$  that is unknown before the propagation of uncertainties [23]. Note that the notation with capital letters is chosen to describe random objects, and the one with small letters still make sense for the deterministic case. Assuming for simplicity that the QoI is a scalar value, with the random input parameters being composed by independent and identically distributed (iid) uniform parameters  $X_i$ , scaled to have support  $[0, 1]^m$ , the Hoeffding-Sobol decomposition is given by

$$Y = \mathcal{M}_0 + \sum_{i=1}^n \mathcal{M}_i(X_i) + \sum_{1 \leq i < j \leq n} \mathcal{M}_{ij}(X_i, X_j) + \dots + \mathcal{M}_{1\dots m}(X_1, \dots, X_n), \quad (1.3)$$

where

$$\begin{aligned} \mathcal{M}_0 &= \mathbb{E}[Y], \\ \mathcal{M}_i(X_i) &= \mathbb{E}[Y | X_i] - \mathcal{M}_0, \\ \mathcal{M}_{ij}(X_i, X_j) &= \mathbb{E}[Y | X_i, X_j] - \mathcal{M}_0 - \mathcal{M}_i - \mathcal{M}_j, \\ &\dots \end{aligned} \quad (1.4)$$

that is,  $\mathcal{M}_0$  is the mean value, and the terms of increasing order are conditional expectations defined in a recursive way, that characterize a unique orthogonal decomposition of the model response [16, 21].

Following this idea, we can now decompose the total variance of the response as follows

$$\text{Var}(Y) = \sum_{\mathbf{u}} \text{Var}(\mathcal{M}_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}})), \quad \text{for } \emptyset \neq \mathbf{u} \subset \{1, \dots, n\}, \quad (1.5)$$

where  $\text{Var}(\mathcal{M}_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}))$  expresses the conditional variance for the subvector  $\mathbf{X}_{\mathbf{u}}$ , containing the variables whose indices are indicated by the subset  $\mathbf{u}$  [21]. Thus, the Sobol' index associated to the subset  $\mathbf{u}$  is defined as the ratio between the contribution given by the interaction among the components of  $\mathbf{u}$  for the model variance, and the total variance itself [14], i.e.,

$$S_{\mathbf{u}} = \frac{\text{Var}(\mathcal{M}_{\mathbf{u}}(\mathbf{X}_{\mathbf{u}}))}{\text{Var}(Y)}. \quad (1.6)$$

As a result of this equation we can verify that, for  $\mathbf{u} \subset \{1, \dots, n\}$ ,  $\mathbf{u} \neq \emptyset$ ,

$$\sum_{\mathbf{u}} S_{\mathbf{u}} = \sum_{i=1}^n S_i + \sum_{1 \leq i < j \leq n} S_{ij} + \dots + S_{1\dots m} = 1, \quad (1.7)$$

that is, by construction the sum of all the Sobol' indices must be equal to the unit.

The terms

$$S_i = \frac{\text{Var}(\mathcal{M}_i(X_i))}{\text{Var}(Y)}, \quad i = 1, \dots, n \quad (1.8)$$

are called the first order Sobol' indices for the single variable  $X_i$  and denote the individual effect of  $X_i$  for the total model variate. Similarly,

$$S_{ij} = \frac{\text{Var}(\mathcal{M}_{ij}(X_{ij}))}{\text{Var}(Y)}, \quad 1 \leq i < j \leq n \quad (1.9)$$

are the second-order indices that contemplate the effect of the interaction between the variables  $X_i$  and  $X_j$ . Keep following we can construct the Sobol' indices of all orders until the  $m$ th order index,  $S_{1,\dots,m}$ , which represents the contribution of the interaction between all the variables in  $\mathbf{X}$  [16].

To measure the full contribution of the  $i$ th random variable  $X_i$  for the total variance either by its single effect or by its interaction with others, we use the total Sobol' indices, which are defined by

$$S_i^T = \sum_{\substack{\mathbf{u} \subset \{1, \dots, n\} \\ i \in \mathbf{u}}} S_{\mathbf{u}} \quad i = 1, \dots, n. \quad (1.10)$$

## 1.4 Surrogate models

To compute Sobol' indices, defined Section 1.3.1, it is necessary to calculate the underlying variances. Despite this task can be done with Monte Carlo (MC) simulation, the associated computational cost can be high (infeasible for high-dimensional systems) and subjected to numerical instabilities like cancellation errors. A very appealing alternative, which allows circumventing these two problems, is the use of surrogate models based on polynomial chaos expansions [15].

### 1.4.1 Polynomial chaos expansion

The Polynomial Chaos Expansion (PCE) of the computational model response is a sum of orthogonal polynomials weighted by coefficients to be determined [22, 23, 25, 26], which reads as

$$Y = \mathcal{M}(\mathbf{X}) = \sum_{\alpha=0}^{\infty} y_{\alpha} \Psi_{\alpha}(\mathbf{X}), \quad (1.11)$$

where  $\Psi_{\alpha}(\mathbf{X})$  are multivariate orthonormal polynomials, associated with the density  $f_{\mathbf{X}}$ , and  $y_{\alpha}$  are the deterministic coefficients to be determined in order to construct the expansion [22]. For computational implementation purposes, a truncated PCE is considered

$$Y \approx \mathcal{M}^{PC}(\mathbf{X}) = \sum_{\alpha=0}^P y_{\alpha} \Psi_{\alpha}(\mathbf{X}), \quad (1.12)$$

where  $P$  is the number of terms in the expansion, which depends on the number of input random variables  $m$  and the maximum degree allowed for the polynomial expansion  $p$ , according to the formula  $P + 1 = (m + p)! / (m! p!)$ . Note the quality of your PCE is directly dependent on the number of terms you have in the expansion [27].

The family of orthonormal polynomials to be used is chosen according to the model input distribution, in a sense that seeks to minimize the number of terms needed in the expansion to build a good computational representation of the model. Table 1.1 summarizes a list of the most classical polynomial families and underlying

distributions. For further details about the construction of the polynomial basis see [22, 23, 27].

**Table 1.1** Correspondence between the random variable distribution and the optimal family of orthonormal polynomials [22].

Type of variable	Distribution	Support	Orthogonal polynomials
Uniform	$\mathbf{1}_{]-1,1[}(x)/2$	$[a, b]$	Legendre
Gaussian	$\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$	$(-\infty, \infty)$	Hermite
Gamma	$x^\alpha e^{-x} \mathbf{1}_{\mathbb{R}_+}(x)$	$[0, \infty)$	Laguerre
Beta	$\mathbf{1}_{]-1,1[}(x) \frac{(1-x)^\alpha (1+x)^\beta}{B(\alpha)B(\beta)}$	$[a,b]$	Jacobi

## 1.4.2 Calculation of the coefficients

Several strategies can be adopted to calculate the PCE coefficients . In this section, we describe the basics of the calculation procedure based on a regression , employing the Ordinary Least-Squares (OLS) method because of its simplicity and generality [24]. The reader is encouraged to see further details about other methods for PCE coefficients calculation in [22, 27].

From the moment that we use the truncated PCE from Eq. (1.11) for computer simulations, there is an error that distances it from the “complete” PCE given by Eq. (1.12), which can be rewritten as follows:

$$Y = \mathcal{M}(\mathbf{X}) = \sum_{\alpha=0}^P y_\alpha \Psi_\alpha(\mathbf{X}) + \varepsilon_P = \mathbf{y}^T \Psi(\mathbf{X}) + \varepsilon_P , \quad (1.13)$$

where  $\varepsilon_P$  is the truncation error ,  $\mathbf{y} = (y_0, \dots, y_P)^T$  is the vector of coefficients and  $\Psi(\mathbf{x}) = \{\Psi_0(\mathbf{x}), \dots, \Psi_P(\mathbf{x})\}$  is the matrix that assembles the values of all the orthonormal polynomials in  $\mathbf{X}$  [27]. Therefore, coefficients of the truncated PCE are calculated in a way to reduce the truncation error [23]. For that we obtain via MC method a “small” set of  $N_s$  samples of each random variable  $X_i$ , called the experimental design

$$\mathcal{X} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N_s)}\} , \quad (1.14)$$

and compute the model response for that samples

$$y^{(1)} = \mathcal{M}(\mathbf{x}^{(1)}), \quad (1.15)$$

$$y^{(2)} = \mathcal{M}(\mathbf{x}^{(1)}), \quad (1.16)$$

$$\vdots \quad (1.17)$$

$$y^{(N_s)} = \mathcal{M}(\mathbf{x}^{(N_s)}). \quad (1.18)$$

The key problem is calculate the PCE coefficients that force the PCE to better fit the responses obtained from the computational model. This is the classic least squares regression problem

$$\mathbf{y}^T \Psi(\mathbf{x}) \approx \mathcal{M}(\mathbf{x}), \quad (1.19)$$

for which the general solution may be expressed as

$$\mathbf{y}^* = \underset{\mathbf{y}}{\operatorname{argmin}} \mathbb{E} \left[ \left( \mathbf{y}^T \Psi(\mathbf{x}) - \mathcal{M}(\mathbf{x}) \right)^2 \right], \quad (1.20)$$

where  $\mathbf{x}$  values comes form the experimental design. Note that MC sampling can be costly, so the idea is to choose  $N_s$  small enough to ensure accuracy without increasing computational cost.

### 1.4.3 Surrogate error estimation

The construction of a good surrogate requires a rigorous process of validation of the response surface obtained. Use a good error metric is essential to characterize a good approximation [15].

Considering the previous section, where PCE coefficients are computed via OLS method, it is natural to evaluate the approximation error by the coefficient of determination [10]. In this case, we can calculate this quantity from the experimental design used in the regression as follows

$$R^2 = 1 - \frac{\frac{1}{N_s} \sum_{i=1}^{N_s} \left( \mathcal{M}(\mathbf{x}^{(i)}) - \mathcal{M}^{PC}(\mathbf{x}^{(i)}) \right)^2}{\hat{V}(Y)}, \quad (1.21)$$

where  $\hat{V}(Y)$  is the empirical variance of the model evaluations, given by

$$\hat{V}(Y) = \frac{1}{N_s - 1} \sum_{i=1}^{N_s} \left( \mathcal{M}(\mathbf{x}^{(i)}) - \bar{y} \right)^2, \quad (1.22)$$

with

$$\bar{y} = \frac{1}{N_s} \sum_{i=1}^{N_s} \mathcal{M}(\mathbf{x}^{(i)}), \quad (1.23)$$

and  $\mathbf{x}^{(i)}$  is the  $i$ th evaluation of the random input  $\mathbf{X}$ . Another option to measure the error is to apply the normalized empirical error [26]. Nevertheless, this measure can be problematic in cases of over-fitting. Thus, if you are not sure about the size of your experimental design, it is recommended to work with the leave-one-out (LOO) cross-validation error [28, 27], calculated by

$$\epsilon_{LOO} = \frac{\sum_{i=1}^{N_s} \left( \mathcal{M}(\mathbf{x}^{(i)}) - \mathcal{M}^{PC \setminus i}(\mathbf{x}^{(i)}) \right)^2}{\sum_{i=1}^{N_s} \left( \mathcal{M}(\mathbf{x}^{(i)}) - \bar{y} \right)^2}, \quad (1.24)$$

where  $\mathcal{M}^{PC \setminus i}$  notation indicates the  $i$ th metamodel built using the reduced experimental design  $\mathcal{X} \setminus \mathbf{x}^{(i)} = \{\mathbf{x}^{(j)}, j = 1, \dots, N_s, j \neq i\}$ .

#### 1.4.4 PCE based Sobol' indices

Note that due the orthonormality of the surrogate PCE metamodel the model variances (partials and total) can be calculated only using the expansion coefficients. The sum of the squared of all the PCE coefficients provides the variance of the model's response and subtracting those coefficients associated with certain indices, the conditioned variances can be obtained [21, 29]. Therefore, an estimator for the Sobol' index associated with the subset  $\mathbf{u}$  is given by

$$S_{\mathbf{u}} = \frac{\sum_{\alpha \in \mathbf{u}} y_{\alpha}^2}{\sum_{\alpha=1}^P y_{\alpha}^2}, \quad (1.25)$$

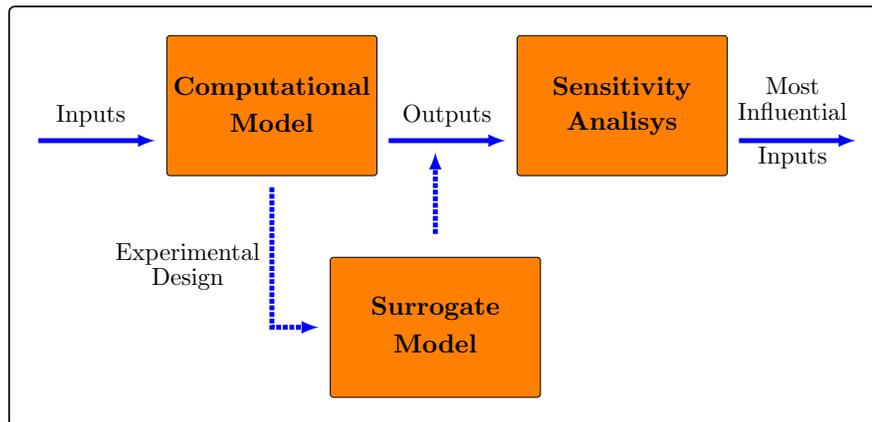
where  $\mathbf{u}$  is a suitable subset of indices.

We observe that once the surrogate is already calculated, these indices can be obtained with negligible computational cost, since in general, a single evaluation of the model is much more expensive than the sums involved in Eq.(1.25) fraction. In addition to the computational cost issue, the use of a surrogate PCE to calculate Sobol' indices also eliminates the possibility of numerical cancellation errors in the variance calculation, since only sums of positive quantities are involved in the above algorithm [15].

## 1.5 A practical tutorial

### 1.5.1 Tutorial description

Finished the presentation of the theoretical background, we now move to the practice, after all, this is a tutorial goal. The computational methodology to compute Sobol' indices is composed of three steps: (i) Characterization of the random input; (ii) Construction of the PCE; (iii) Calculation of the Sobol' indices . The framework of this process is illustrated in the Fig. 1.2.



**Fig. 1.2** Schematic representation of the process of obtaining the Sobol' indices by using a surrogate model .

Here UQLab [30] library is explored to facilitate the implementation of a code to compute the Sobol' indices . We encourage the reader to look for more details in the UQLab manual [30] . To describe the random input it is necessary to define the distribution for each random parameter as well as the respective hyperparameters and support. In the second step, you have to define a computational model to simulate your computational experiment, the maximum degree of your polynomials expansion and the number of samples that will be used to generate the design set for the regression . The PCE coefficients are calculated for each time of observation defined, and you have everything you need to obtain the Sobol' indices up to the desired order. Note that you can not select the maximum order for the Sobol' indices higher than the dimension of the random input vector. Further details about this construction can be seen in UQLab manual for Global Sensitivity Analysis [29].

### 1.5.2 SoBioS: Sobol' indices for Biological Systems

To facilitate the manipulation of the UQLab packages , the authors developed a computational library called SoBioS – Sobol' indices for Biological Systems – focused on the simulation of the Sobol' indices for biological systems with the UQLab tool. This computational library is available in the following repository:

<https://americocunhajr.github.io/SoBioS>

The auxiliary routines used to create the following example results can be also found there, but, in resume, you need two basic routines: The main file to define the input details, to call the computational model, to perform the sensitivity analysis and plot results, and a QoI file, to define your quantity of interest.

### 1.5.3 Example 1: predator-prey dynamics

The first example is the classical Lotka-Volterra model . Better known as predator-prey model , this is a dynamical system used to reproduce a simple predator-prey relationship [18]. It is assumed that preys are capable of reproduce spontaneously and the predators are only able to feed from the previous prey. The standard model consider preys and predators' reproduction, the natural death of the predator (independent of the prey), and death of preys by hunting. So, the relation between the two species is beneficial for the predator and harmful for the prey.

By assuming a simple representation of this mechanisms, we can construct the following pair of equations

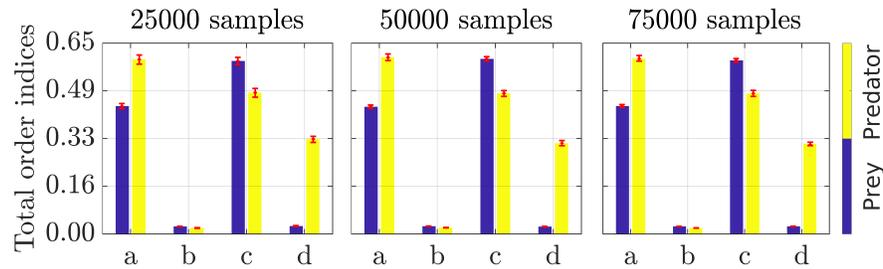
$$\frac{dV}{dt} = aV - bVP , \quad (1.26)$$

$$\frac{dP}{dt} = dVP - cP , \quad (1.27)$$

where  $V$  and  $P$  represent the populations of prey and predator, respectively, and the time is measured in years. In the first equation we have the subtraction between the birth term and the hunt term. Before the hunt, is assumed that the predator increases in a proportion of the energetic efficiency  $d$  times the number of preys, and decreases with a constant ratio. The model parameters as well as the other simulation parameters are described in the Table 1.2. The model parameters' intervals and initial conditions were extracted from the UQLab Bayesian Calibration manual [31] and the other simulation values were chosen by the authors. Numerical integration were performed using the Dormand-Prince adaptive Runge-Kutta method, implemented in ode45, and the response will be restricted to three times instants to simplify the visualization of the results. Our interest can be for the prey or predator population depending of the situation.

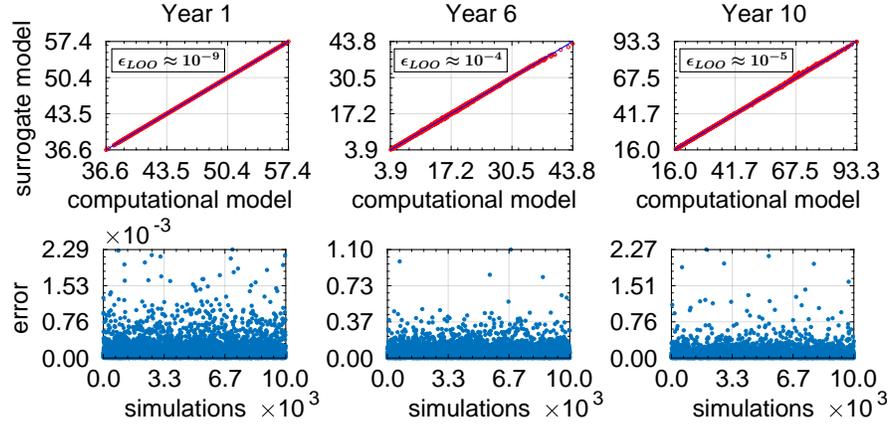
**Table 1.2** Parameters description and values used to simulate the Sobol' indices for the predator-prey model

Name	Description	Value
$a$	birth rate for preys	$[0.44, 0.68]$ year <sup>-1</sup>
$b$	search rate	$[0.02, 0.044]$ year <sup>-1</sup>
$c$	death rate for predators	$[0.71, 1.15]$ year <sup>-1</sup>
$d$	energetic efficiency	$[0.0226, 0.0354]$ year <sup>-1</sup>
$V_0$	initial quantity of preys	33 individuals
$P_0$	initial quantity of predators	6.2 individuals
tspan	time span	$[0, 10]$ years
tQoI	time instants of interest	$\{1, 6, 10\}$ years
QoI	$V$ or $P$	$V(t)$ or $P(t)$ individuals

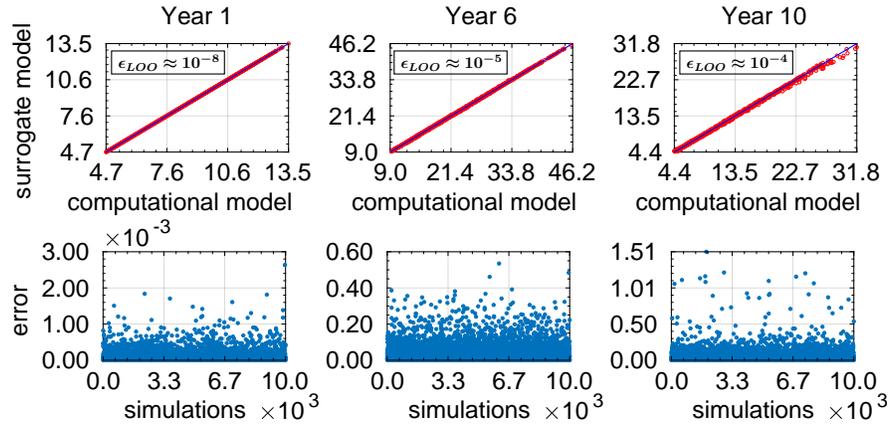
**Fig. 1.3** Comparison between MC-Sobol' Total Order results using 25000 (left), 50000 (middle) and 75000 (right) samples with 95% confidential intervals (red) calculated with Bootstrap method considering both predators (yellow) and prey (blue) populations in the 10<sup>th</sup> year.

In this first example we compare the results of sensitivity analysis by MC and Surrogate approaches. Before that, it is necessary to analyze how much samples to use for each one of those methods. Two different strategies are considered in order to evaluate the reliability of the computed Sobol' indices. For the MC method, the idea is to execute hierarchical simulations increasing the number of samples, using only one value of each QoI, and calculating 95% confidence intervals (CI) by Bootstrap method [23]. For PCE-Sobol the approach is to plot the validation graphs and compare the surrogate response with the full model one, to estimate also the ideal maximum polynomial degree. Some results for each strategy can be observed in Figures 1.3 – 1.5, respectively. We can see the convergence as we increase the number of samples in MC because the CI amplitude decreases. On the validation plots of Figures 1.4 and 1.5 we can observe the good match between the surrogate results and the response for the original computational model in addition to the low order for the  $\epsilon_{LOO}$  cross validation error. This comparison is done using 10000 samples. By these results we define a size of 75000 samples for MC simulations and

1000 for PCE simulations. Additionally, the ideal maximum degree for the surrogate approximation was adopted as 6.



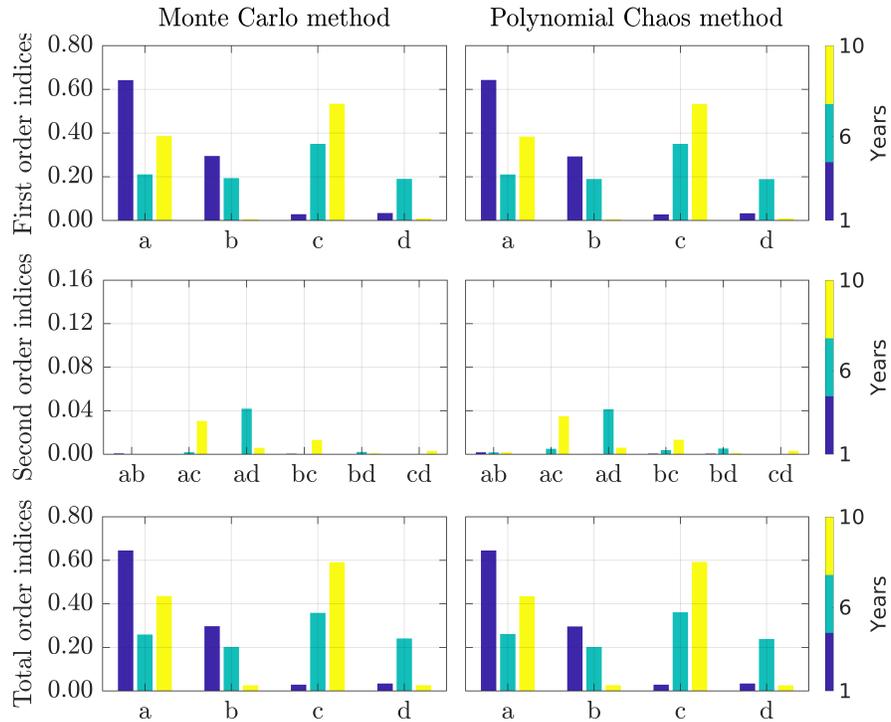
**Fig. 1.4** Comparison PCE surrogate response and computational model response (top) for predators in years 1 (left), 6 (middle) and 10 (right), associated mismatch error (bottom), and leave-one-out error (top legend) using 1000 samples as experimental design and 10000 as validation set.



**Fig. 1.5** Comparison PCE surrogate response and computational model response (top) for preys in years 1 (left), 6 (middle) and 10 (right), associated mismatch error (bottom), and leave-one-out error (top legend) using 1000 samples as experimental design and 10000 as validation set.

Finally, Figures 1.6 and 1.7 present the comparison of the MC and PCE results for sensitivity analysis. In these graphs, we can observe that the results are pretty close. However, the total order for MC is slightly smaller. This is due to the negative second-order indices obtained since it suffers from a cancellation error.

, something to which the calculation via PCE is immune once it does not involve subtractions. This fact has been omitted from the associated figures to avoid any confusion about negative Sobol' indices , which do not make sense.

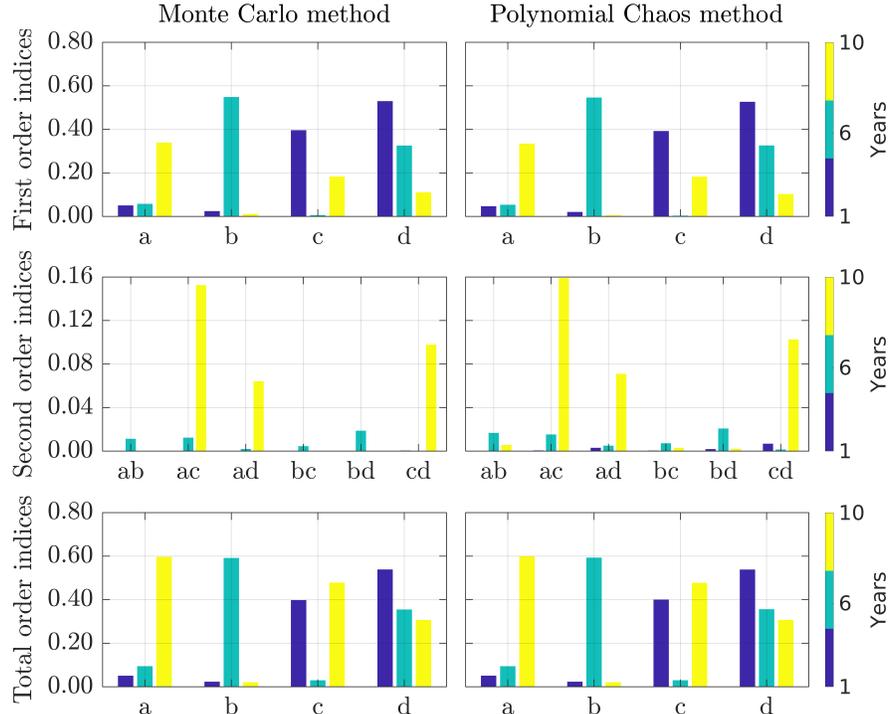


**Fig. 1.6** Comparison between MC-Sobol' results (left) and PCE-Sobol' results (right) for the prey population of the Lotka-Volterra model . The simulation was performed with 75000 samples for the MC and 1000 for the PCE .

Note that for this simple scenario the cancellation error is not that problematic but wondering a situation of high order indices, the error propagation can affect your results' conclusions. The second advantage is, of course, in the time of simulation. Even for this simple analysis, the difference is notable. Using an Intel Core i5-8250U 1.6Ghz x 8 (8GB of RAM), the MC method needed 2 minutes, while the PCE performed in 45s, even with the second one plotting more results due to the validation graphs.

Besides all that discussion about methods, the Sobol' results reveal that, for this parametric intervals and initial conditions, the parameter  $d$  is more important in the first year,  $b$  in the sixth and  $a$  in the final one when analysing the predator population. For prey the results is completely different. The parameter  $a$  control the most part of the variance in the first week but the  $c$  parameters assume the control after that.

The great difference of scenarios for each population can sound weird as well as the change of the most important parameter during the passage of time. The truth is



**Fig. 1.7** Comparison between MC-Sobol' results (left) and PCE-Sobol' results (right) for the prey population of the Lotka-Volterra model . The simulation was perform with 75000 samples for the MC and 1000 for the PCE .

that is not only possible but common. This detail for systems with time dependency is essential for control measures. It will be necessary some adaptive strategy to affect different parameters of phenomena at each step of implementing the measure. Also, note that in this case the second-order interactions were not so much important for the total order indices in case of prey but increases significantly the total indices for predators.

### 1.5.4 Example 2: NF- $\kappa$ B signalling pathway

NF- $\kappa$ B is a family of transcription factors (TF) that takes part is the regulation of several mammalian cellular processes, to mention some: cell division, apoptosis, inflammation, immune response and cancer disease [32]. The family consists of five subunits (p50, p52, p65, c-Rel and RelB), which associate to form functional dimers. The NF- $\kappa$ B complex when bound to the I $\kappa$ B inhibitor protein is inactive in the cytoplasm, being released after the phosphorylation of the I $\kappa$ B protein by the I $\kappa$ B kinase (IKK) complex. This dissociation of the NF- $\kappa$ B:I $\kappa$ B complex also

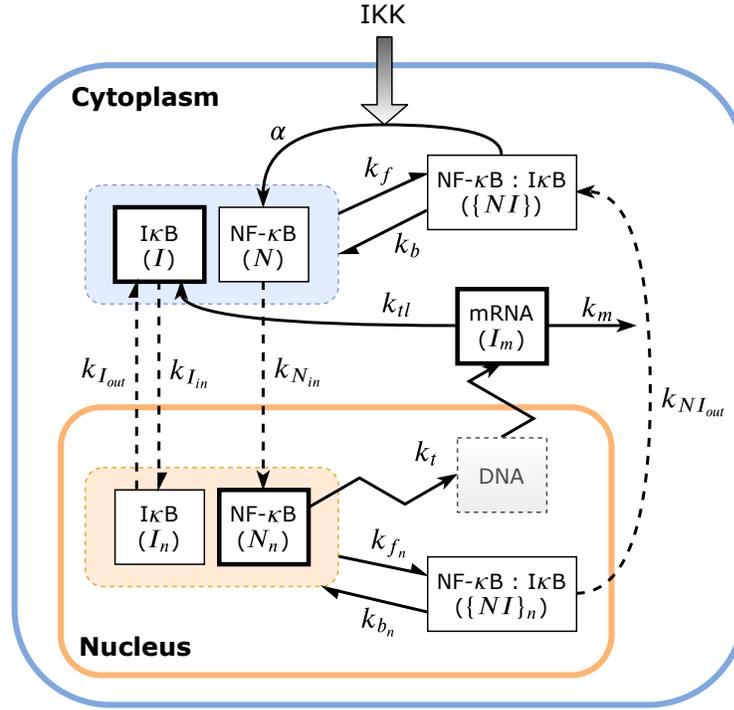
promotes the  $I\kappa B$  ubiquitination and than its degradation. At this moment,  $NF-\kappa B$  is free to be translocated into the nucleus, where it activates gene transcription by binding to specific DNA  $\kappa B$  sites. Among the activated ones are the genes for  $I\kappa B$ , that is,  $I\kappa B$  mRNA are transcribed and translated into new  $I\kappa B$  proteins. Part of it binds to the  $NF-\kappa B$  in the cytoplasm, and part enters the nucleus where it binds to nuclear  $NF-\kappa B$  forming a  $NF-\kappa B:I\kappa B$  that is exported to the cytoplasm. In both cases, the complex is again the target of the  $I\kappa B$  kinase (IKK), characterizing a negative feedback mechanism.

A mathematical model for the  $NF-\kappa B$  pathway was introduced in 2002 by Hoffman et al. [5]. Their model consists of 26 molecular species (variables) and 64 reaction coefficients (parameters). Besides setting up the system of differential equations for such a complex biological system, they studied the influence of the negative feedback on the sustained and on the damped oscillation of the  $NF-\kappa B$  concentration over time. An interesting follow up of their group's work was presented in [6]. It is clear from the references list how their mathematical model was influential and how the  $NF-\kappa B$  pathway model is a well succeed system biology example of the interplay between mathematical modeling and experimental analyses.

Taking the bi-compartmental model of [5] as departure point, removing species that have no feedback from  $NF-\kappa B$  and removing slow reactions at the expense of faster ones, Krishna et al. [33] were able to extract the core feedback loop of the model, coming up with a reduced model constituted of 7 species (variables) and 12 reaction coefficients (parameters). The variables of the reduced model are:  $N_n$  and  $N$ , the free nuclear and free cytoplasmic  $NF-\kappa B$  concentration, respectively;  $I_n$  and  $I$ , the free nuclear and free cytoplasmic  $I\kappa B$  concentration, respectively;  $I_m$ , the  $I\kappa B$  mRNA concentration;  $\{NI\}_n$  and  $\{NI\}$ , the nuclear and cytoplasmic  $NF-\kappa B:I\kappa B$  complex concentration, respectively; and  $IKK$ , the  $I\kappa B$  kinase concentration. The dynamics of the seven species are depicted in Fig. 1.8, resulting in the system of equations

$$\begin{aligned}
\frac{dN_n}{dt} &= k_{N_{in}}N - k_{f_n}N_nI_n + k_{b_n}\{NI\}_n, \\
\frac{dN}{dt} &= -k_{N_{in}}N - k_fNI + (k_b + \alpha)\{NI\}, \\
\frac{dI_n}{dt} &= k_{I_{in}}I - k_{I_{out}}I_n - k_{f_n}N_nI_n + k_{b_n}\{NI\}_n, \\
\frac{dI}{dt} &= -k_{I_{in}}I + k_{I_{out}}I_n - k_fNI + k_b\{NI\} + k_{tl}I_m, \\
\frac{d\{NI\}}{dt} &= k_{NI_{out}}\{NI\}_n + k_fNI - (k_b + \alpha)\{NI\}, \\
\frac{d\{NI\}_n}{dt} &= -k_{NI_{out}}\{NI\}_n + k_{f_n}N_nI_n - k_{b_n}\{NI\}_n, \\
\frac{dI_m}{dt} &= k_tN_n^2 - k_mI_m.
\end{aligned} \tag{1.28}$$

The description and the nominal values of the parameters are shown in Table 1.3, according to [33]. In the model, the NF- $\kappa$ B pathway is activated by an extracellular stimuli, considered by a I $\kappa$ B kinase (IKK) input, incorporated in the parameter  $\alpha$  (top of Fig. 1.8).



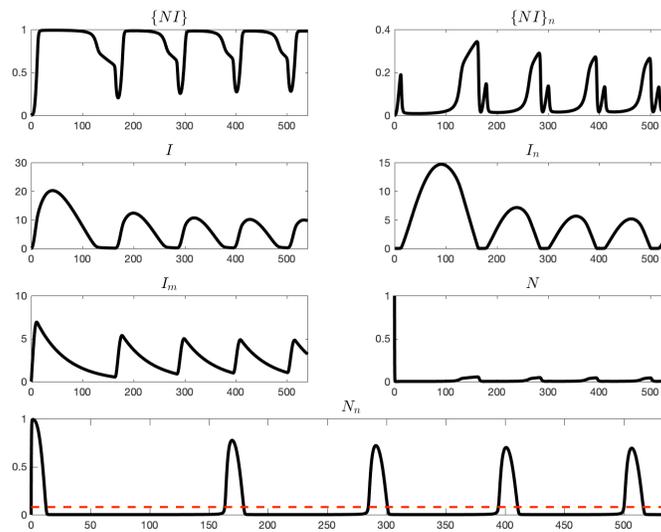
**Fig. 1.8** Diagram representing the dynamics of the seven species NF- $\kappa$ B model [33]. The parameters are described in Table 1.3. The bold boxes are the species present on the reduced three species model.

Before proceeding we actually note that for I $\kappa$ B concentration is meant I $\kappa$ B $\alpha$  concentration, since as elucidated in [5] is the I $\kappa$ B $\alpha$  isoform of the I $\kappa$ B protein that enters in the negative feedback loop of the NF- $\kappa$ B pathway, and generates a sustained oscillation on the NF- $\kappa$ B concentration. The solution of the system for the nominal values in Table 1.3 is shown in Fig. 1.9, with initial conditions  $(N_n, N, I_n, I, \{NI\}, \{NI\}_n, I_m)_0 = (0, 1, 0, 0, 0, 0, 0)$  and  $IKK = 0.7$ , and it was obtained by numerically integrating the system (1.28) with the ODE solver, `ode15s`, for stiff equations, where it is observed the spiky oscillation of the NF- $\kappa$ B concentration in time. With further considerations on the NF- $\kappa$ B:I $\kappa$ B complex association and dissociation reactions, they reduced the model even more, coming up with an over reduced model constituted of 3 species:  $N_n, I, I_m$  (the bold boxes in Fig. 1.8),

**Table 1.3** Parameters descriptions and nominal values of the seven species NF- $\kappa$ B model [33]

Parameter	Values (Units)	Description
$k_{Nin}$	$5.4 \text{ min}^{-1}$	NF- $\kappa$ B nuclear import rate
$k_{Iin}$	$0.018 \text{ min}^{-1}$	I $\kappa$ B nuclear import rate
$k_{Iout}$	$0.012 \text{ min}^{-1}$	I $\kappa$ B cytoplasm export rate
$k_{NIout}$	$0.83 \text{ min}^{-1}$	NF- $\kappa$ B:I $\kappa$ B cytoplasm export rate
$k_f$	$30 \mu\text{M}^{-1} \cdot \text{min}^{-1}$	NF- $\kappa$ B:I $\kappa$ B cytoplasm association rate
$k_{fn}$	$30 \mu\text{M}^{-1} \cdot \text{min}^{-1}$	NF- $\kappa$ B:I $\kappa$ B nuclear association rate
$k_b$	$0.03 \text{ min}^{-1}$	NF- $\kappa$ B:I $\kappa$ B cytoplasm dissociation rate
$k_{bn}$	$0.03 \text{ min}^{-1}$	NF- $\kappa$ B:I $\kappa$ B nuclear dissociation rate
$k_{tl}$	$0.24 \text{ min}^{-1}$	I $\kappa$ B mRNA translation rate
$k_t$	$1.03 \mu\text{M}^{-1} \cdot \text{min}^{-1}$	I $\kappa$ B mRNA transcription rate
$k_m$	$0.017 \text{ min}^{-1}$	mRNA degradation rate
$\alpha$	$1.05 \times IKK \text{ min}^{-1}$	I $\kappa$ B degradation rate (by phosphorylation)

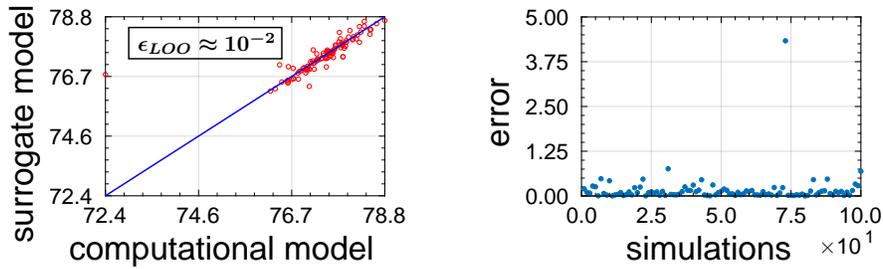
and 5 parameters (here they are not straight reaction coefficients as before, but combinations of those).



**Fig. 1.9** Numerically integrated solution for the system (1.28), where it can be observed the spiky oscillation of the NF- $\kappa$ B concentration in time. The constant dashed (red) line is  $N_n = \text{const.} = \text{mean}(N_n)$ .

As pointed out in [33], the spiky type oscillations are robust with respect to the  $IKK$  input variation and that the down-regulated genes by NF- $\kappa$ B have different times

of response, related to this oscillation, that is why the QoI for our sensitivity analysis was the peak duration, that was defined in [33] as the time the NF- $\kappa$ B concentration spends above its mean. For the solution in Fig. 1.9, the mean is represented by a constant dashed (red) line in the last (bottom) subplot. To build the PCE surrogate model 400 experimental design samples were taken with a maximum polynomial degree of 12. The probabilistic input model consisted of independent uniform random variables for each of the parameters, with lower and upper bounds given by 1.5% of dispersion around the mean, taken as the nominal values considered at Table 1.3. The quality of the surrogate approximation is appreciated in Fig. 1.10, showing to be reasonable to the purpose. According to the total indices the most five influential parameters, in decreasing order, are:  $k_t, k_{tI}, k_{Iin}, k_m, k_{Nin}$ . It is interesting to note that these parameters are the ones that enter in the over reduced model composed of three species, this is because these parameters appear in the equations of (1.28) related to the species  $N_n, I, I_m$ . The transcription and the translation parameters of the I $\kappa$ B mRNA are very influential as can be noted from the first order indices, Fig. 1.12 and its degradation rate  $k_m$  does not have such a first order influence, but  $k_m$  appears with significant coupling with five other species in the second order level, Fig. 1.13. Actually, the three highest second order indices, namely,  $k_{Nin} - k_m, k_t - k_m$  and  $k_{Iin} - k_m$  are coupled with it. Another parameter that have no first order influence, but appears coupled with five other species in the second order level, is the parameter  $k_{Nin}$ . That is why both  $k_m$  and  $k_{Nin}$  express themselves in the total order indices but not in the first order ones. Finally we need to mention that in Fig. 1.13, we are showing only the second order indices above the threshold value  $10^{-4}$ .



**Fig. 1.10** Comparison PCE surrogate response and original computational model response for the peak duration QoI.

### 1.5.5 Example 3: The SIR model

The famous SIR model is a classical tool to study epidemics [1]. Recently several research groups are using variations of the SIR model in the new coronavirus pandemic research [9]. Of course this model, as will show below, is not ideal to reproduces the

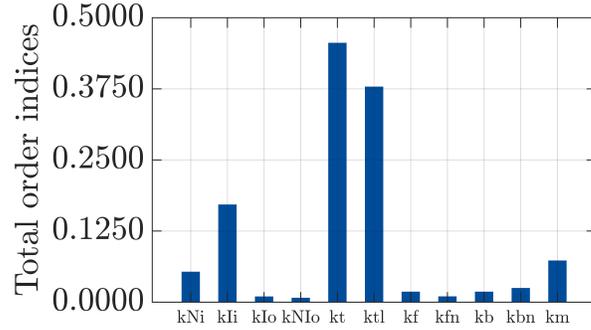


Fig. 1.11 Total order Sobol' indices for the peak duration QoI.

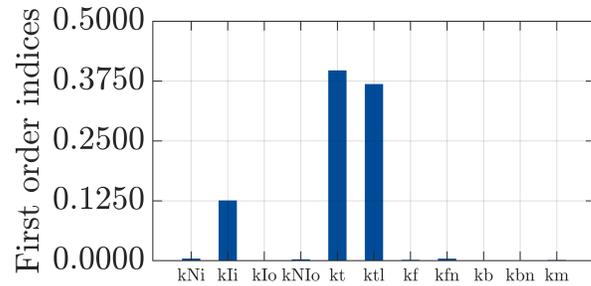


Fig. 1.12 First order Sobol' indices for the peak duration QoI.

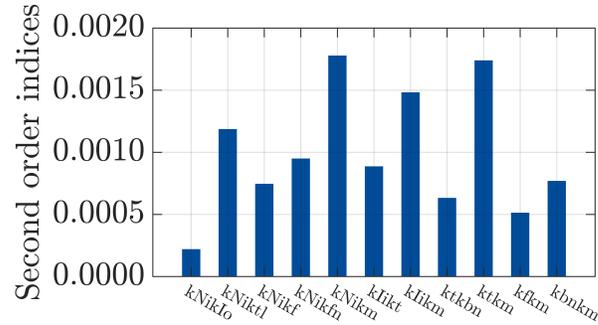


Fig. 1.13 Second order Sobol' indices for the peak duration QoI.

Corona phenomena correctly, however can be useful to test hypotheses and analyse scenarios. The model is based on separating the host population from a disease in three compartments: The Susceptible,  $S$ , are the population of individuals able to become infected by the disease upon contact with the pathogen; Infected,  $I$ , indicates those who carry the pathogen and have the ability of infect the susceptible; The

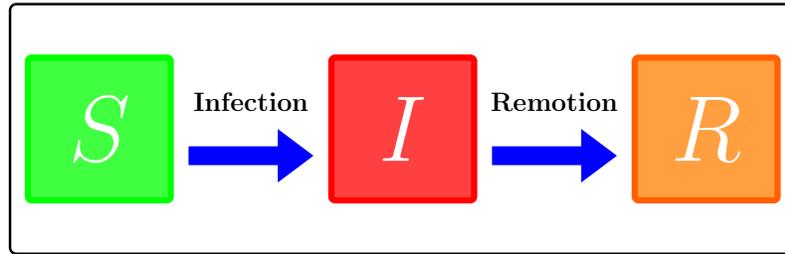
Removed,  $R$ , reunites the individuals who recovered from the disease (gaining immunity) or died from it [1]. The evolution between the compartments is represented in Figure 1.14 and mathematically formulated by the following set of equations:

$$\frac{dS}{dt} = -\frac{bSI}{N}, \quad (1.29)$$

$$\frac{dI}{dt} = \frac{bSI}{N} - aI, \quad (1.30)$$

$$\frac{dR}{dt} = aI, \quad (1.31)$$

where  $bSI/N$  is the infections rate term and  $aI$  indicates the removal rate. The QoI in this model is the infected population  $I(t)$ . The description and the values of the simulation parameters are gathered in Table 1.4. As in the previous example, we have reference for the mean values of the random variables, and the dispersion factor around the nominal values will be of 0.4 for each one. The parameter  $a$  can be interpreted as the inverse of the infectious period. Thereby, we took the mean value of 6.5 days estimated by the researchers of the Imperial College group (in their 9th Report) [9]. Besides that, in the same report it is estimated the basic number of reproduction  $\mathcal{R}_0 = b/a$  for several countries. Using this equation of  $\mathcal{R}_0$  for the SIR model [34] and  $a = 1$ , we estimate a mean for the transmission rate.



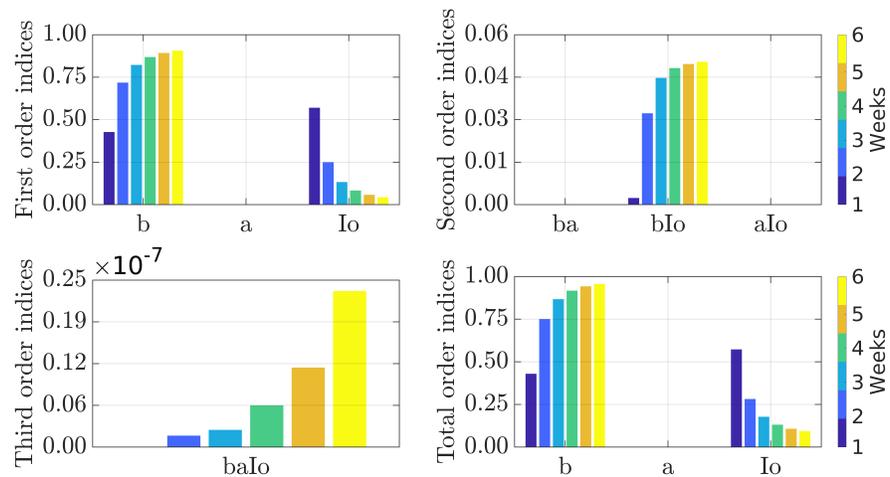
**Fig. 1.14** Schematic representation of the SIR compartmental model's mechanisms.

Different of the previous examples, now we will consider also a initial condition as a random variable, to account for the uncertainty in the initial number of infected individuals, that is not well known. If we assume that the host population is constant, the initial number of susceptible is also a random variable, depending on the initial number of infected. Finally, although our system is measured in days, we will display results in weeks to facilitate visualization.

In the Figures 1.15 and 1.16 the reader can find the Sobol' indices results and validation plots, respectively. The maximum degree for the PCE was 10 and the experimental design was compose of 1000 samples. It can be noted a very good match between the computational response and the surrogate. To differentiate of the previous two examples, here we calculate the third order Sobol' index. In this scenario this index do not show so much information, but reveal that the influence of this third

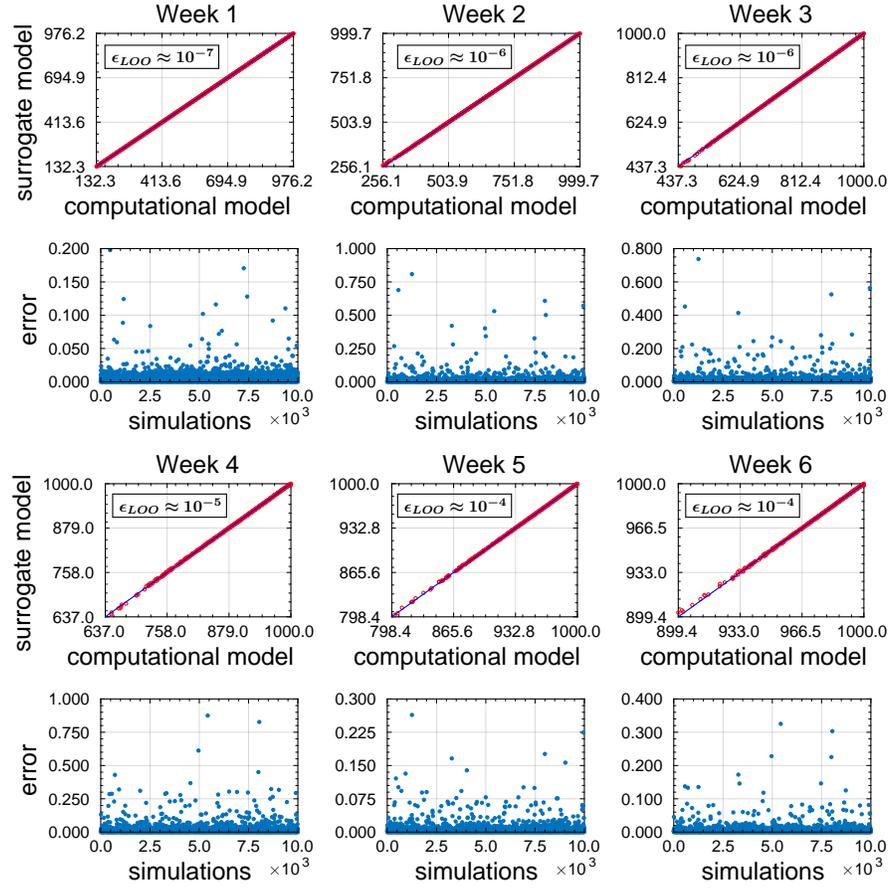
**Table 1.4** Parameters description and values used to simulate the Sobol' indices for the SIR model

Name	Description	Value
$b$	transmission rate	$0.37 \text{ days}^{-1} \text{ individuals}^{-1}$
$a$	removal rate	$0.15 \text{ days}^{-1}$
$N$	total population	1000 individuals
$R_0$	initial number of removed	0 individuals
$I_0$	initial number of infected	200 individuals
$S_0$	initial number of susceptible	$(N - I_0)$ individuals
tspan	time span	$[0, 42]$ days
tQoI	time instants of interest	$\{7, 14, \dots, 42\}$ days
QoI	$I$	$I(t)$ individuals



**Fig. 1.15** Sobol' indices for the infected population of the SIR model in the weeks of interest simulated using 1000 samples.

order term maybe increases over time. The same seems occur with the  $b$  importance while the opposite can be observed for  $I_0$ . So, maybe if we analyse a previous time windows, the initial number of infected individuals can be more influence for the sensitivity of the response. Similarly, for a posterior time window we could have a  $I_0$  not important. Of course, new studies are need to support these hypothesis but is quite interesting to imagine these possibilities. Curiously, the removal rate had practically zero importance. This is not intuitive because we expected some kind of influence over time. But, it is important to be clear that this not mean that the  $a$  parameter not change that response if it changes. This result showed that if the two parameters and the initial number of infected individuals changes,  $b$  and  $I_0$  will domain how the response will change.



**Fig. 1.16** Comparison PCE surrogate response and original computational model response for infected population in weeks of interest (red curves), and associated leave-one-out error (legends) using 1000 samples as experimental design and 10000 as validation set.

## 1.6 Final remarks

This chapter presented the use of Sobol' indices for global sensitivity analysis in biological systems. This type of formalism allows us to statistically measure how each of the system's parameters contributes to its response. This tool can be very useful to identify which factors have a greater contribution in a certain quantity of interest associated with the biological system, which can be interesting, for instance, in the calibration process of a computational model concerning a certain set of observations (data). This analysis tool is presented in the form of a tutorial, with a simplistic summary of the theory, without much mathematical formalism in favor of a better understanding of the fundamental ideas, and application in three biological systems with different levels of complexity. A code in MATLAB was developed and

is available in a public repository to facilitate the use of this tool in the most diverse types of biological systems .

## Acknowledgments

The authors would like to thank Prof. Flávio Codeço Coelho (EMAp/FGV) by discussing several ideas at the beginning of the preparation of this manuscript. They are also grateful, for the financial support provided, to Carlos Chagas Filho Research Foundation of Rio de Janeiro State (FAPERJ) under grants 210.021/2018 and 211.037/2019, and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

## References

1. F. Brauer, *Infectious Disease Modelling* **2**(2), 113 (2017). DOI 10.1016/j.idm.2017.02.001
2. H.W. Hethcote, *SIAM Review* **42**(4), 599 (2000). DOI 10.1137/S0036144500371907
3. D.S. Boukal, A. Bideault, B.M. Carreira, A. Sentis, *Current Opinion in Insect Science* **35**, 88 (2019). DOI 10.1016/j.cois.2019.06.014
4. Y. Sekerci, S. Petrovskii, *Bulletin of Mathematical Biology* **77**, 2325–2353 (2015). DOI 10.1007/s11538-015-0126-0
5. A. Hoffmann, A. Levchenko, M.L. Scott, D. Baltimore, *Science* **298**(5596), 1241 (2002). DOI 10.1126/science.1071914
6. R. Cheong, A. Hoffmann, A. Levchenko, *Molecular Systems Biology* **4**(1), 192 (2008). DOI 10.1038/msb.2008.30
7. Centers for Disease Control and Prevention. Preventing the spread of coronavirus disease 2019 in homes and residential communities. <https://www.cdc.gov/coronavirus/2019-ncov/hcp/guidance-prevent-spread.html> (2020). Accessed: 03/20/20
8. World Health Organization. Coronavirus disease (COVID-19) outbreak. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019> (2020). Accessed: 03/19/20
9. I.C.C..R. Team, Covid-19 reports. Tech. Rep. 1-14, Imperial College, <https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/> (2020)
10. H. Motulsky, A. Christopoulos, *Fitting Models to Biological Data Using Linear and Nonlinear Regression: A Practical Guide to Curve Fitting*, 1st edn. (Oxford University Press, Oxford, 2004)
11. E. Dantas, M. Tosin, A. Cunha Jr, *Applied Mathematics and Computation* **338**, 249 (2018). DOI 10.1016/j.amc.2018.06.024
12. E. Dantas, M. Tosin, A. Cunha Jr, in *Proceedings of the Conference of Computational Interdisciplinary Science* (2019)
13. A. Saltelli, et al., *Environmental Modelling & Software* **114**, 29 (2019). DOI 10.1016/j.envsoft.2019.01.012
14. R.C. Smith, *Uncertainty Quantification*, 1st edn. (SIAM, Philadelphia, 2014)
15. L.L. Gratiet, S. Marelli, B. Sudret, *Metamodel-Based Sensitivity Analysis: Polynomial Chaos Expansions and Gaussian Processes* (Springer, 2017), pp. 1289–1326
16. I.M. Sobol, *Mathematical Modelling in Civil Engineering* **55**(1-3), 407 (2001)
17. J.T. Oden, *An Introduction to Mathematical Modeling: A Course in Mechanics*, 1st edn. (Wiley, New Jersey, 2011)

18. J.W. Haefner, P. Spanos, E. Akin, *An Introduction to Mathematical Epidemiology*, 1st edn. (Springer, New York, 2015)
19. C. Soize, *Uncertainty Quantification: An Accelerated Course with Advanced Applications in Computational Engineering*, 1st edn. (Springer, New York, 2017)
20. J. Müller, C. Kuttler, *Methods and Models in Mathematical Biology: Deterministic and Stochastic Approaches*, 1st edn. (Springer, New York, 2015)
21. K. Konakli, B. Sudret, *Computer Methods in Applied Mechanics and Engineering* **263**, 42 (2016). DOI 10.1016/j.res.2016.07.012
22. D. Xiu, *Numerical Methods for Stochastic Computations: A Spectral Method Approach*, 1st edn. (Princeton University Press, New Jersey, 2010)
23. R. Ghanem, D. Higdon, H. Owhadi (eds.), *Handbook of Uncertainty Quantification*, 1st edn. (Springer, New York, 2017)
24. B. Sudret, *Reliability Engineering and System Safety* **93**(7), 964 (2008). DOI 10.1016/j.res.2007.04.002
25. R.G. Ghanem, P.D. Spanos, *Stochastic Finite Element Method: A Spectral Approach*, revised edn. (Dover Publications, 2003)
26. B. Sudret, G. Blatman, M. Berveiller, *Response Surfaces based on Polynomial Chaos Expansions* (ISTE/Wiley, 2013), pp. 147–168. DOI 10.1002/9781118601099.ch8
27. S. Marelli, B. Sudret, UQLab user manual – Polynomial chaos expansions. Tech. rep., Chair of Risk, Safety & Uncertainty Quantification, ETH Zurich (2018). Report # UQLab-V1.1-104
28. G. Blatman, B. Sudret, *Probabilistic Engineering Mechanics* **25**(2), 183 (2010). DOI 10.1016/j.probenmech.2009.10.003
29. S. Marelli, C. Lamas, B. Sudret, K. Konakli, C. Mylonas, UQLab user manual – Sensitivity analysis. Tech. rep., Chair of Risk, Safety & Uncertainty Quantification, ETH Zurich (2018). Report # UQLab-V1.1-106
30. S. Marelli, B. Sudret, in *Proc. 2nd Int. Conf. on Vulnerability, Risk Analysis and Management (ICVRAM2014)*, Liverpool, United Kingdom (2014), p. 2554–2563. DOI 10.1061/9780784413609.257
31. P.R. Wagner, J. Nagel, S. Marelli, B. Sudret, UQLab user manual – Bayesian inversion for model calibration and validation. Tech. rep., Chair of Risk, Safety & Uncertainty Quantification, ETH Zurich (2019). Report # UQLab-V1.2-113
32. B.R.B. Pires, A.L. Mencalha, G.M. Ferreira, W.F. de Souza, J.A. Morgado-Díaz, A.M. Maia, S. Corrêa, E.S.F.W. Abdelhay, *PLOS ONE* **12**(1), 1 (2017). DOI 10.1371/journal.pone.0169622
33. S. Krishna, M.H. Jensen, K. Sneppen, *Proceedings of the National Academy of Sciences* **103**(29), 10840 (2006). DOI 10.1073/pnas.0604085103
34. M. Martcheva, *An Introduction to Mathematical Epidemiology*, 1st edn. (Springer, New York, 2015)