

Article

# Quantifying Parameter Interdependence in Stochastic Discrete Models of Biochemical Systems

Samaneh Gholami \* and Silvana Ilie 

Department of Mathematics, Toronto Metropolitan University, Toronto, ON M5B 2K3, Canada; silvana@torontomu.ca

\* Correspondence: samaneh.gholami@torontomu.ca

**Abstract:** Stochastic modeling of biochemical processes at the cellular level has been the subject of intense research in recent years. The Chemical Master Equation is a broadly utilized stochastic discrete model of such processes. Numerous important biochemical systems consist of many species subject to many reactions. As a result, their mathematical models depend on many parameters. In applications, some of the model parameters may be unknown, so their values need to be estimated from the experimental data. However, the problem of parameter value inference can be quite challenging, especially in the stochastic setting. To estimate accurately the values of a subset of parameters, the system should be sensitive with respect to variations in each of these parameters and they should not be correlated. In this paper, we propose a technique for detecting collinearity among models' parameters and we apply this method for selecting subsets of parameters that can be estimated from the available data. The analysis relies on finite-difference sensitivity estimations and the singular value decomposition of the sensitivity matrix. We illustrated the advantages of the proposed method by successfully testing it on several models of biochemical systems of practical interest.

**Keywords:** stochastic simulation algorithm; stochastic biochemical systems; sensitivity analysis; finite-difference methods; parameter subset selection; estimability analysis



**Citation:** Gholami, S.; Ilie, S. Quantifying Parameter Interdependence in Stochastic Discrete Models of Biochemical Systems. *Entropy* **2023**, *25*, 1168. <https://doi.org/10.3390/e25081168>

Academic Editor: Pavel Kraikivski

Received: 3 July 2023

Revised: 31 July 2023

Accepted: 3 August 2023

Published: 5 August 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Mathematical and computational modeling have become widespread in the study of complex dynamical systems, particularly in investigating cellular processes and biochemical networks [1]. Frequently, mathematical modeling of chemical reaction systems relies on deterministic differential equations and mass action kinetics. However, biochemical systems in the cell are intrinsically noisy [2,3], and thus stochastic models must be employed to account for the random fluctuations observed experimentally, especially when some species have low molecular counts [4,5]. One of the most popular stochastic discrete models of biochemically reacting systems is the Chemical Master Equation [6,7]. This model is utilized to describe the dynamics of systems for which molecular populations of some species are low or noise is significant. It assumes that the system state is a Markov process [6]. It is generally impracticable to solve this model analytically, except for very simple systems.

Gillespie developed the Stochastic Simulation Algorithm (SSA) [8,9], a Monte Carlo technique for simulating statistically exact realizations of the stochastic process whose distribution is governed by the Chemical Master Equation. The random time change representation of the stochastic process depicting the system state was introduced in [10]. Based on this representation, Rathinam et al. [11] designed an exact Monte Carlo method for the Chemical Master Equation, the Random Time Change algorithm. Other simulation strategies for stochastic models of biochemically reacting systems were presented in the literature (for references see, e.g., [12–15]).

The biochemical networks arising in applications may be quite complex, involving many reactions and/or species, which means that their mathematical models have many

parameters. Some of the values of a model's kinetic parameters may not be known [16,17] and they may need to be estimated from the available data. Also, certain parameters have a substantial influence on the system's output. Thus, it is essential to study the system's behavior when these parameters are perturbed. While stochastic discrete models of biochemical systems capture the inherent randomness observed in cellular processes, they pose challenges with regard to their parameter estimation and identification. Hence, developing efficient and accurate methods for identifying and estimating their parameters would be a key advance in studying these models.

Practical identifiability (or estimability) analysis aims to establish if the parameters can be accurately and reliably estimated from the available data [18]. In this context, identifiable parameters are those which can be determined with high confidence from the observed behavior of the system; otherwise, the parameters are unidentifiable. Using practical identifiability, one can select subsets of parameters that significantly impact the behavior of the system. If the parameters in such a subset are not interdependent, then they are identifiable. These parameters can be accurately estimated when sufficient and quality data is available, and their accurate estimation is crucial for building the model. Also, these parameters may provide insight into the key underlying mechanisms of the biochemical system. Furthermore, the identifiability analysis helps select the unidentifiable parameters, which have a negligible impact on the model behavior and can be eliminated, thus guiding model reduction. There exist numerous studies of identifiability analysis for deterministic models, such as the reaction rate equations [19–26]. Nonetheless, much less work has been dedicated to parameter estimability of stochastic models of biological processes.

One important method for practical identifiability is to utilize sensitivity analysis. Local sensitivity analysis assesses the change in the system's behavior caused by a small variation in the value of a certain parameter. Insignificant changes in the system dynamics indicate that the specific parameter is not important, and thus it is not identifiable. Also, a parameter is not identifiable if it is correlated with other parameters, such that a variation in its value can be compensated by suitable adjustments in other parameters. For stochastic models, finite-difference methods can be used to estimate the sensitivity of the expected value of the given function of the system state. In the class of finite-difference sensitivity estimators for the Chemical Master Equation, those employing exact Monte Carlo simulation methods are the Coupled Finite Difference method of Anderson [27], the Common Reaction Path scheme (based on the Random Time Change algorithm) and the Common Random Number strategy (utilizing the SSA) of Rathinam et al. [11]. These estimators utilize coupled perturbed and unperturbed trajectories to approximate sensitivities. The coupling lowers the variance of the estimator so that the method requires fewer realizations to achieve the same accuracy of the estimation. Due to this, the computational time of the algorithm is reduced, for a prescribed accuracy. Of the three strategies, the Coupled Finite Difference algorithm has the lowest variance of the estimator [28]. These schemes perform best for non-stiff models. For stiff problems, finite-difference techniques can be applied with various coupled tau-leaping strategies to increase the efficiency of the simulation [29].

In this work, we consider the problem of practical parameter identifiability for stochastic discrete biochemical networks modeled with the Chemical Master Equation. This is a critical problem, and a direct extension of the techniques developed for ordinary differential equations to stochastic discrete models is not possible. Our contribution is generalizing a method by Gábor et al. [30] to find the highest parameter identifiable sets for models of biochemical systems, from the continuous deterministic to the stochastic discrete models of well-stirred biochemical systems, which is a difficult task. The proposed method identifies the subsets of parameters that are independent and significant for the model's behavior, based on the existing data, and thus are identifiable. We utilize local sensitivity estimations to study parameter estimability. For approximating sensitivities, we apply finite-difference techniques, namely the Coupled Finite Difference [27], the Common Reaction Path, and the Common Random Number methods [11]. We make use of the normalized sensitivity matrix to develop several identifiability metrics, which adapt existing techniques for the

reaction rate equations [19,20] to the more challenging Chemical Master Equation model. In addition, we apply the singular value decomposition of the non-dimensional sensitivity matrix, to determine its rank. This analysis helps gain insight into the interrelations between parameters. Furthermore, the proposed methodology can be employed to decide which parameters can be reliably estimated from the available data, for the Chemical Master Equation, and may assist experimental design for more accurate parameter approximations. It is worth noting that, in general, the expected value of the system state governed by the Chemical Master Equation may not satisfy the deterministic reaction rate equations, when some reactions are of second or higher order [14].

This paper is structured as follows. Section 2 is dedicated to the background on stochastic discrete models for well-stirred biochemical networks and their simulation methods, parametric sensitivity schemes for stochastic and deterministic models, and practical identifiability techniques, including the new algorithm for selecting subsets of identifiable parameters. The proposed algorithm is tested on various stochastic models arising in applications in Section 3. Section 4 presents a summary of our results.

## 2. Materials and Methods

### 2.1. Background

Suppose a system has  $N$  biochemical species, denoted by  $S_1, S_2, \dots, S_N$ , that undergo  $M$  distinct chemical reactions. It is maintained at a constant temperature, in a constant volume. Provided that the biochemical network is well-stirred, it may be represented by a state vector,

$$X(t) = [X_1(t), X_2(t), \dots, X_N(t)]^T,$$

where  $X(t)$  has entries  $X_i(t)$ , the amount of  $S_i$  molecules in the system at time  $t$ . A reaction  $R_j$  produces a variation in the system state, which is given by the state change vector  $v_j \in \mathbb{R}^N$ ,

$$v_j = [v_{1j}, v_{2j}, \dots, v_{Nj}]^T,$$

where  $v_{ij}$  is the perturbation in the molecular amount of  $S_i$  after the reaction fires. If one reaction  $R_j$  happens during the time interval  $[t, t + \Delta t]$ , then the resulting state is  $X(t + \Delta t) = X(t) + v_j$ . The array having  $v_j$  as the  $j$ -th column is called the stoichiometric matrix. Also associated with the reaction  $R_j$ , we can define the propensity function  $a_j$ , by  $a_j(x)dt =$  the probability that a single reaction  $R_j$  occurs between  $[t, t + dt)$ , assuming that the system state at time  $t$  is  $x$ . The form of the propensity function  $a_j$  is determined by the type of reaction. For a first-order reaction,  $S_m \xrightarrow{c_j} \text{products}$ , the propensity is expressed as  $a_j(X(t)) = c_j X_m(t)$ . For a second-order reaction,  $S_m + S_n \xrightarrow{c_j} \text{products}$ , the propensity is  $a_j(X(t)) = c_j X_m(t) X_n(t)$ , if  $m \neq n$  and  $a_j(X(t)) = \frac{1}{2} c_j X_m(t) (X_m(t) - 1)$ , if  $m = n$ .

#### 2.1.1. Chemical Master Equation

To study the behavior of the well-stirred biochemical system, we need to determine  $P(x, t|x_0, t_0)$ , the probability of the system state being  $X(t) = x$  at time  $t$ , if at  $t_0$  it was  $X(t_0) = x_0$ . This probability satisfies the Chemical Master Equation [6,7]

$$\frac{d}{dt}P(x, t|x_0, t_0) = \sum_{j=1}^M [a_j(x - v_j)P(x - v_j, t|x_0, t_0) - a_j(x)P(x, t|x_0, t_0)]. \tag{1}$$

This is a stochastic discrete model. It is a linear system of ordinary differential equations, each equation describing the probability of the system being in a particular state  $x$ . The biochemical system state  $X(t)$  is a discrete in space and continuous in time Markov process. The space of all possible states is typically quite large, and in such cases the Chemical Master Equation is of very high dimension. Therefore, it is challenging to solve it directly, except for some simple systems.

As an alternative to solving the Chemical Master Equation directly, it is possible to generate correct trajectories one by one. Gillespie [8,9] proposed a Monte Carlo strategy to compute such trajectories, which are in exact agreement with the probability distribution associated with the discrete stochastic model (1). The strategy, also referred to as the Stochastic Simulation Algorithm (SSA), has been broadly employed for solving stochastic models in Systems Biology [3,14,31]. The SSA is described below.

#### Gillespie's Algorithm

1. Initialize the time  $t \leftarrow t_0$  and the state of the system,  $X(t) \leftarrow x_0$ .
2. While  $t < T$
3. Calculate each propensity  $a_j(X(t))$  for  $j = 1, \dots, M$  and the sum  $a_0(X(t)) \leftarrow \sum_{r=1}^M a_r(X(t))$
4. Sample two uniform random variables over  $[0, 1]$ , to obtain  $\eta_1, \eta_2$ .
5. Evaluate the time  $\tau$  and the index  $j$  of the next occurring reaction, according to
  - (a)  $\tau \leftarrow -(\ln \eta_1) / a_0(x)$
  - (b)  $j \leftarrow$  the smallest integer fulfilling  $\sum_{r=1}^j a_r(x) > \eta_2 a_0(x)$
6. Update the state  $X(t + \tau) \leftarrow X(t) + v_j$  and the time  $t \leftarrow t + \tau$ .
7. End while.

The Random Time Change (RTC) algorithm [11], based on the Random Time Change representation [10], is another exact Monte Carlo simulation strategy for the Chemical Master Equation. We refer the reader to [11] for details on this algorithm.

#### 2.1.2. Chemical Langevin Equation

An intermediate model between the Chemical Master Equation and the reaction rate equation is the Chemical Langevin Equation [32]. This is a system of stochastic differential equations of size equal to the number of reacting species. The Chemical Langevin Equation is a reduction in the Chemical Master Equation model assuming that the biochemical system has a macroscopically infinitesimal scale in time step such that, over  $\delta t$ , every reaction occurs multiple times and, at the same time, its propensity function does not vary significantly. Under these assumptions, the system state is governed by

$$dX(t, c) = \sum_{j=1}^M v_j a_j(X(t, c), c) dt + \sum_{j=1}^M v_j \sqrt{a_j(X(t, c), c)} dW_j(t) \quad (2)$$

where  $W_j$  are independent Wiener processes for  $j = 1, \dots, M$ . The state  $X(t)$  may be approximated by a Markov process continuous in space. Equation (2) represents the Chemical Langevin Equation.

#### 2.1.3. Reaction Rate Equation

A coarser level of resolution in modeling biochemically reacting networks is provided by the continuous deterministic model of chemical kinetics. This model, known as the reaction rate equations, is valid under the assumption of the thermodynamic limit. In the thermodynamic limit, the molecular amounts for all species and the system volume tend towards infinity, as the concentrations of species within the system remain constant. Hence, the stochastic terms in the Chemical Langevin Equation are much smaller than the deterministic terms. As a result, the Chemical Langevin Equation model reduces to the reaction rate equations, in the thermodynamic limit. This condition is satisfied when all  $S_i$  molecular counts are very large. The reaction rate equations (RRE) are of the form

$$\frac{dX(t, c)}{dt} = \sum_{j=1}^M v_j a_j(X(t, c), c). \quad (3)$$

Equation (3) is a set of ordinary differential equations, with one equation for each biochemical species. In the event that all reactions in the system are of order at most one, the reaction rate equation can be obtained from the Chemical Master Equation (1), by considering the expected value of the system state. However, in general, the evolution of the mean trajectory in the Chemical Master Equation model does not obey the continuous deterministic model. Then, the RRE does not properly depict the true behavior of the biochemical network. In fact, there are numerous cellular networks for which noise significantly influences the system dynamics [12,31,33].

## 2.2. Parametric Correlations

Sensitivity analysis plays a central role in constructing models [24]. It assesses how changes in parameters cause variations in a model's output. If a negligible adjustment in a parameter leads to significant alterations in the outcomes, we consider the model to be sensitive to that specific parameter. Precise estimations are not necessary for parameters with low sensitivity. Conversely, parameters with high associated sensitivity become key control points for the behavior of the system. In what follows, we shall focus on the sensitivity analysis of system outputs with respect to rate parameters.

### 2.2.1. Parametric Sensitivity for the Chemical Master Equation

Let  $f$  be a function of interest of the system state and  $c$  a model parameter. In the stochastic setting, the local sensitivity with respect to a parameter  $c$  is defined as  $\frac{\partial}{\partial c} \mathbb{E}[f(X(t, c))]$  where  $\mathbb{E}(\cdot)$  is the expected value. Popular methods for estimating local sensitivities with respect to the model's parameters for the Chemical Master Equation often rely on finite-difference schemes and Monte Carlo methods for generating the perturbed and unperturbed trajectories. By forward finite-difference schemes, one can estimate  $\frac{\partial}{\partial c} \mathbb{E}[f(X(t, c))] \approx \{\mathbb{E}[f(X(t, c + \theta))] - \mathbb{E}[f(X(t, c))]\} / \theta$ , where  $\theta$  is a small perturbation of the parameter of interest,  $c$ . To efficiently approximate the sensitivity by Monte Carlo methods, the trajectories for  $X(t, c + \theta)$  and  $X(t, c)$  are generated using common random numbers. Among such methods are the Common Random Number (CRN), the Common Reaction Path (CRP) algorithms [11], and the Coupled Finite-Difference (CFD) algorithm [27].

### 2.2.2. Common Random Number

The Common Random Number presented in [11] is a finite-difference numerical method for estimating parametric sensitivities for the stochastic discrete model (1). It reuses random numbers to generate the perturbed and unperturbed paths. In doing so, it reduces the variance of the sensitivity estimator, and thus it has increased efficiency compared to a strategy based on independent random numbers. For the  $r$ -th iteration, it computes two SSA trajectories,  $X_{[r]}(t, c + \theta)$  -the perturbed and  $X_{[r]}(t, c)$  -the unperturbed path, each employing the same stream of uniform  $(0, 1)$  random numbers. Usually, the coupling of the CRN technique is less efficient than that of the CRN and CFD schemes [27]. The sensitivity of the  $r$ -th path is approximated by

$$Z_{[r]}(t, c) = \frac{f(X_{[r]}(t, c + \theta)) - f(X_{[r]}(t, c))}{\theta}, \quad (4)$$

while an estimate of the sensitivity is obtained from the sample mean  $(\sum_{i=1}^R Z_{[r]}(t, c)) / R$ ,  $R$  being the number of paired trajectories simulated.

### 2.2.3. Common Reaction Path

The Common Reaction Path technique is also a finite-difference sensitivity estimator for the Chemical Master Equation [11]. The CRP strategy applies the RTC algorithm to sim-

ulate sample paths. In this method, coupling of the processes involves some independent unit-rate Poisson processes,  $\{Y_j\}_{1 \leq j \leq M}$ . The coupling of the perturbed— $X(\cdot, c + \theta)$  and unperturbed— $X(\cdot, c)$  processes is achieved using the random time change representation

$$\begin{aligned}
 X(t, c) &= x_0 + \sum_{j=1}^M v_j Y_j \left( \int_0^t a_j(X(s, c), c) ds \right) \\
 X(t, c + \theta) &= x_0 + \sum_{j=1}^M v_j Y_j \left( \int_0^t a_j(X(s, c + \theta), c + \theta) ds \right)
 \end{aligned}
 \tag{5}$$

The  $r$ -th iteration of the CRP algorithm generates the paired trajectories  $X_{[r]}(t, c + \theta)$  and  $X_{[r]}(t, c)$  with the RTC algorithm, each using the same  $M$  independent streams of unit-rate exponential random numbers. As before, the sensitivity of the  $r$ -th trajectory is estimated by (4). This coupling has been shown to be typically stronger than that of the CRN method, leading to a lower variance of the estimation [11,27].

#### 2.2.4. Coupled Finite-Difference

Another finite-difference sensitivity estimator for the stochastic discrete model is the Coupled Finite-Difference scheme [27]. The CFD method relies on the random time change representation of the unperturbed and perturbed processes

$$\begin{aligned}
 X(t, c) &= x_0 + \sum_{j=1}^M v_j Y_j^{(1)} \left( \int_0^t \min(a_j(X(s, c), c), a_j(X(s, c + \theta), c + \theta)) ds \right) \\
 &+ \sum_{j=1}^M v_j Y_j^{(2)} \left( \int_0^t [a_j(X(s, c), c) - \min(a_j(X(s, c), c), a_j(X(s, c + \theta), c + \theta))] ds \right) \\
 X(t, c + \theta) &= x_0 + \sum_{j=1}^M v_j Y_j^{(1)} \left( \int_0^t [\min(a_j(X(s, c), c), a_j(X(s, c + \theta), c + \theta))] ds \right) \\
 &+ \sum_{j=1}^M v_j Y_j^{(3)} \left( \int_0^t [a_j(X(s, c + \theta), c + \theta) - \min(a_j(X(s, c), c), a_j(X(s, c + \theta), c + \theta))] ds \right)
 \end{aligned}
 \tag{6}$$

where  $\{Y_j^{(1)}\}_{1 \leq j \leq M}$ ,  $\{Y_j^{(2)}\}_{1 \leq j \leq M}$ , and  $\{Y_j^{(3)}\}_{1 \leq j \leq M}$  are independent unit-rate Poisson processes. Furthermore, the CFD strategy uses a version of the Next Reaction Method to compute the coupled perturbed and unperturbed trajectories,  $X_{[r]}(t, c + \theta)$  and  $X_{[r]}(t, c)$ , and (4) to approximate the local sensitivity of the  $r$ -th path. Among the finite-difference sensitivity estimators with exact underlying simulation techniques for the CME, the CFD performs the best, followed by the CRP and the CRN [27,28]. Indeed, the CFD achieves the smallest variance of the sensitivity estimator of the three methods described above [28]. As a consequence, for the same number of trajectories simulated, we shall consider in our investigations the CFD sensitivity approximations to be the most accurate and reliable.

#### 2.2.5. Parametric Sensitivity for the Chemical Langevin Equations

Glasserman [34] developed a technique for computing pathwise parametric sensitivities for certain problems modeled by stochastic differential equations. This method was applied to the Chemical Langevin Equation (CLE) model in [33]. For computing the sensitivity of each path, we differentiate Equation (2) with respect to parameter  $c$  and obtain

$$\begin{aligned}
 d\left(\frac{\partial X}{\partial c}\right) &= \sum_{j=1}^M v_j \left[ \frac{\partial a_j(X)}{\partial X} \frac{\partial X}{\partial c} + \frac{\partial a_j(X)}{\partial c} \right] (t) dt \\
 &+ \sum_{j=1}^M v_j \left[ \frac{1}{2\sqrt{a_j(X)}} \left( \frac{\partial a_j(X)}{\partial X} \frac{\partial X}{\partial c} + \frac{\partial a_j(X)}{\partial c} \right) \right] (t) dW_j.
 \end{aligned}
 \tag{7}$$

Solving the coupled system of Equations (2) and (7) for  $(X, \partial X/\partial c)$  will determine the pathwise sensitivities. At time  $t = 0$ , the local sensitivities with respect to the rate parameters

are zero. The Chemical Langevin Equation is, in general, valid when all molecular amounts are sufficiently large. Effective simulation strategies for this model require adaptive time-stepping methods [35,36].

### 2.2.6. Parametric Sensitivity for the Reaction Rate Equations

In the deterministic scenario, the behavior of the biochemical system is governed by the reaction rate Equation (3). To find the local sensitivity for this model, the derivative with respect to the desired kinetic parameter is applied to Equation (3), yielding

$$\frac{d}{dt} \mathcal{S} = \sum_{j=1}^M v_j \left( \frac{\partial a_j(X(t,c),c)}{\partial c} + \sum_{i=1}^N \frac{\partial a_j(X(t,c),c)}{\partial X_i} \mathcal{S}_i \right). \quad (8)$$

Here,  $\mathcal{S} = \partial X(t,c)/\partial c$  is the sensitivity with respect to parameter  $c$ . The sensitivity is computed by solving for  $(X, \mathcal{S})$  the system of ordinary differential Equations (3) and (8), with the initial conditions  $X(0,c) = x_0$  and  $\mathcal{S}(0) = 0$ . The deterministic model is applicable when all reacting molecular populations are very large. Nonetheless, when low molecular counts of some species exist or noise plays a significant role, this approach may fail in accurately capturing the characteristics of the biochemical system. Then, deterministic techniques for sensitivity-based identifiability analysis are not valid.

### 2.3. Practical Identifiability Analysis

When a model's performance is investigated, it is important to evaluate the accuracy of the parameter values. Still, poor or noisy data, interdependence of parameters, or weak dependence of the system dynamics on certain parameters may hinder the accurate estimation of parameter values. As a result, it is possible for these values to change significantly, without influencing the model's output. Consequently, the concept of identifiability is essential for the analysis of a mathematical model [19,24].

Identifiability can be classified into two main categories: structural identifiability and practical identifiability. For a structurally identifiable model, there exists a unique parameterization for any specified output of the model (see, e.g., [21,26]). On the other hand, practical identifiability involves detecting non-identifiable parameters by fitting the model to data that closely resemble the available observations (see, e.g., [18,19,22,25] for analyses of deterministic models). For this type of identifiability, it is helpful to study the parametric sensitivity of the model. In this work, we use sensitivity-based identifiability for the Chemical Master Equation. We determine identifiability and collinearity indexes by generalizing methods for deterministic models [19] to the more challenging case of stochastic discrete biochemical systems.

#### 2.3.1. Sensitivity-Based Identifiability Analysis

Several identifiability strategies for deterministic models exist in the literature. One such approach by Brun et al. [19] is based on local sensitivity analysis of deterministic models. Sensitivity analysis quantifies the impact of parameter variations on the system's dynamics.

Below, we review some techniques for identifiability analysis of deterministic models relying on local parametric sensitivity. These techniques can be applied to the reaction rate Equation (3). Denote by

$$S_{ik}(X, t, c) = \frac{\partial X_i(t, c)}{\partial c_k} \quad (9)$$

the local sensitivity of the molecular amount  $X_i(t, c)$  at time  $t$ , with respect to the kinetic parameter  $c_k$ . For time  $t$ , the parametric sensitivity matrix is  $S(X, t, c) = \frac{\partial}{\partial c} X(t, c) =$

$\{S_{ik}(X, t, c)\}_{1 \leq i \leq N, 1 \leq k \leq M}$ . In addition, the non-dimensional sensitivity coefficient corresponding to the  $i$ -th species and the parameter  $c_k$  at time  $t$  is

$$s_{ik}(t) = \frac{c_k}{X_i(t, c)} \frac{\partial X_i(t, c)}{\partial c_k}. \tag{10}$$

Here,  $c = [c_1, \dots, c_M]$  is the vector of kinetic parameters associated to reactions  $\{R_j\}_{1 \leq j \leq M}$ . Furthermore, let  $t_1 < t_2 < \dots < t_L$  be a sequence of time-points spanning the integration interval  $[0, T]$ . Ideally, some of these time-points should be inside the interval corresponding to the biochemical network’s transient behavior, when applicable. Also, consider the concatenated non-dimensional sensitivity matrix, for all the time-points in the grid, and apply the normalization (10) for each entry,

$$s(X, c) = \begin{pmatrix} s_{11}(t_1) & \cdots & s_{1M}(t_1) \\ \vdots & \ddots & \vdots \\ s_{N1}(t_L) & \cdots & s_{NM}(t_L) \end{pmatrix}. \tag{11}$$

To rank the parameters of the model, we utilize the non-dimensional sensitivity matrix of size  $(NL) \times M$  from (11). The  $k$ -th column in this matrix measures the sensitivities with respect to  $c_k$ , the rate parameter of reaction  $R_k$ . Let us calculate the norm of each column in the sensitivity matrix (11) to obtain a parameter ranking. The norm of each column  $s_k(X, c) = [s_{1k}(t_1), \dots, s_{Nk}(t_1), \dots, s_{1k}(t_L), \dots, s_{Nk}(t_L)]^T$  serves as a measure of the significance of parameter  $c_k$  on the dynamics of the system. A higher norm indicates that altering that parameter value has a substantial impact on the system state. Parameters can be arranged in order of their significance. The following sensitivity measure is employed for evaluating the significance of the parameters, based on the sensitivity matrix (adapted after [19])

$$\delta_k^{msqr} = \sqrt{\frac{1}{n} \sum_{i=1}^n s_{ik}^2}. \tag{12}$$

The larger the measure  $\delta_k^{msqr}$ , the more significant the parameter  $c_k$  is (for  $1 \leq k \leq M$ ).

### 2.3.2. Parameter Collinearity

Extensive research has been conducted to examine the collinearity in various problems. Brun et al. [19] introduces a strategy for identifying parameter relationships based on collinearity analysis, in the deterministic framework, and presents a novel approach to explore the connections between parameters. Note that the columns of a matrix  $B$  are nearly linearly dependent (or near collinear) if a non-zero vector  $z = [z_1, \dots, z_M]^T$  exists such that  $Bz \approx 0$ , where  $B$  has  $M$  columns. If the  $Bz = 0$  holds and  $z \neq 0$ , the columns of  $B$  are linearly dependent (or collinear).

Now, take the normalized sensitivity matrix  $\tilde{S}$ , having as the  $m$ -th column the vector

$$\tilde{s}_m(X, c) = \frac{s_m(X, c)}{\|s_m(X, c)\|_2},$$

for  $1 \leq m \leq M$ . It is useful to first normalize these vectors, to prevent biases due to differences in the absolute value of local sensitivities for various parameters. A large norm of  $\|s_m\|_2$  indicates that a small variation in parameter  $c_m$  can significantly impact the system’s behavior; thus, this parameter is important. For this parameter to be identifiable, it should not be correlated with other parameters.

Let us consider any subsets  $K$  of  $k$  parameters ( $k \leq M$ ) from the set of parameters  $\{c_1, c_2, \dots, c_M\}$  and the corresponding sub-matrix  $\tilde{S}_K(X, c)$  of the normalized sensitivity

matrix, with columns the  $k$  sensitivity vectors. A measure of collinearity of the subset  $K$  of parameters, with corresponding matrix  $\tilde{S}_K$ , is given by

$$CI_K = \frac{1}{\min_{\|z\|_2=1} \|\tilde{S}_K z\|_2} = \frac{1}{\sqrt{\lambda_k}} \quad (13)$$

where  $\lambda_k$  is the minimum eigenvalue of the matrix  $\tilde{S}_K^T \tilde{S}_K$  and  $\|\cdot\|_2$  is the norm-2 of a vector. The measure (13) is known as the collinearity index of the subset  $K$  [19,30]. The closest the columns of the matrix  $\tilde{S}_K$  are to a linearly dependent set of vectors, the smallest  $\min_{\|z\|_2=1} \|\tilde{S}_K z\|_2$  is. Thus, a large collinearity index  $CI_K$  indicates a high level of collinearity of the parameters in the set. This implies that changes in the model dynamics due to small perturbations in one of the parameters of the almost collinear set may be prevented by suitable variations in the other parameters of the set. As a consequence, if a set of parameters is collinear, it is not identifiable. According to [19], a subset of parameters is considered identifiable if the associated collinearity index satisfies  $CI_K < 20$ . With this observation, it is possible to uncover the subsets of model parameters that can be identified as well as those that cannot be identified. The collinearity index may be computed for all the subsets  $K$  of the parameter space, to determine the parameter subsets that are not collinear. When a group of parameters has a high collinearity index, any set containing it as a subset will also have a high collinearity index.

Another technique to assess the identifiability of the model parameters is to use the singular value decomposition (SVD) of a matrix. In general, the SVD [37,38] of an  $n \times M$  matrix  $s$  is

$$s = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \quad (14)$$

where the  $\mathbf{U}$  is an  $n \times n$  unitary matrix,  $\mathbf{V}$  is an  $M \times M$  unitary matrix and  $\mathbf{\Sigma}$  is an  $n \times M$  non-negative diagonal matrix with the diagonal entries

$$\sigma_1 \geq \sigma_2 \geq \dots \sigma_r > \sigma_{r+1} = \dots = \sigma_M = 0.$$

The values  $\{\sigma_m^2\}_{1 \leq m \leq M}$  are the eigenvalues of the matrix  $s^T s$ . The index  $r$  measures the rank of the matrix  $s$  and it is the largest number of linearly independent columns of this matrix. Numerically, the singular values  $\sigma_{r+1}, \dots, \sigma_M$ , which are below a specified small tolerance are considered practically zero. In this work, we use the singular value decomposition of the matrix  $s$  to determine its rank. This rank is a reliable measure of the number of rate parameters that are not collinear. Furthermore, zero or very close to zero singular values show that the group of all the reaction rate parameters of the model are collinear. Therefore, there are some model parameters that cannot be estimated from the available data.

Brun et al. [20] also introduced a determinant measure

$$\rho_k = \det(S_K^T S_K)^{1/2k} \quad (15)$$

to find the appropriate number of parameters to estimate.

The metrics considered above can be utilized to determine the identifiability of parameter sets as follows. The sensitivity measure  $\delta_k^{msqr}$  is used to evaluate the importance of each parameter  $c_k$ . On the other hand, the collinearity index measures whether the set  $K$  of parameters are independent, whenever  $CI_K < 20$ . In the case that both conditions are satisfied, (a) the parameters in the subset  $K$  are not collinear and (b) each parameter in the group is important, the parameters in  $K$  are identifiable. Finally, the determinant  $\rho_K$  can be employed to compare the identifiability of various groups of parameters.

### 2.3.3. Method for Selecting Subsets of Identifiable Parameters

The practical identifiability methods presented above were developed for continuous deterministic models [19,20], and are thus applicable for the reaction rate equation model. However, this model may fail to faithfully represent the behavior of biochemical systems,

which involve low molecular counts of some species. Consequently, new methodologies are required for the parameter identifiability of stochastic discrete models of biochemical systems. In this work, we develop novel strategies for determining sets of identifiable parameters for the Chemical Master Equation. We generalize the work of Gábor et al. [30] on identifying subsets of identifiable parameters in deterministic models, to address the much more challenging case of stochastic discrete models of well-stirred biochemical systems. This generalization is essential as stochasticity plays a significant role in accurately modeling real-world biological systems, and our approach allows for an in-depth study of more complex biochemical networks encountered in applications.

The measures presented above were designed for deterministic models. We aim to adapt these measures to systems modeled by the Chemical Master Equation. For this model, the sensitivity coefficients are computed as

$$S_{ik}(\mathbb{E}[X], t) = \frac{\partial}{\partial c_k} \mathbb{E}[X_i(t, c)].$$

Then, we shall compute the sensitivity matrix for the CME according to

$$S(t) = \frac{\partial \mathbb{E}[X(t, c)]}{\partial c} = \begin{pmatrix} \frac{\partial}{\partial c_1} \mathbb{E}(X_1(t, c)) & \cdots & \frac{\partial}{\partial c_M} \mathbb{E}(X_1(t, c)) \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial c_1} \mathbb{E}(X_N(t, c)) & \cdots & \frac{\partial}{\partial c_M} \mathbb{E}(X_N(t, c)) \end{pmatrix}. \tag{16}$$

Take a sequence of time-points  $0 = t_1 < t_2 < \dots < t_L = T$ , relevant to the biochemical system under consideration. The fully normalized (non-dimensional) sensitivity coefficient of the  $i$ -th species with respect to the  $c_k$  parameter at time  $t_\ell$  is

$$s_{ik}(t_\ell) = \frac{c_k}{\mathbb{E}[X_i(t_\ell, c)]} \frac{\partial}{\partial c_k} \mathbb{E}[X_i(t_\ell, c)] \quad \text{for } 1 \leq i \leq N, 1 \leq k \leq M. \tag{17}$$

The concatenated non-dimensional sensitivity matrix over these discrete time-points with entries (17) is

$$s(\mathbb{E}[X], c) = \begin{pmatrix} s_{11}(t_1) & \cdots & s_{1M}(t_1) \\ \vdots & \ddots & \vdots \\ s_{N1}(t_L) & \cdots & s_{NM}(t_L) \end{pmatrix}. \tag{18}$$

Normalizing the  $\ell$ -th column of matrix (18), namely  $s_\ell(\mathbb{E}[X], c)$ , gives

$$\tilde{s}_\ell(\mathbb{E}[X], c) = \frac{s_\ell(\mathbb{E}[X], c)}{\|s_\ell(\mathbb{E}[X], c)\|_2}. \tag{19}$$

Finally, the normalized sensitivity matrix  $\tilde{S}$  has  $\tilde{s}_\ell(\mathbb{E}[X], c)$  as its  $\ell$ -th column. For the Chemical Master Equation, the sensitivity measure  $\delta_k^{msqr}$  and the collinearity index  $CI_K$  are computed using (12) and (13), respectively, for the sensitivity matrix of the expected value  $\mathbb{E}[X]$  rather than the system state  $X$ , as was the case for the reaction rate equation.

Moreover, we will employ the finite-difference methods described above to estimate parametric sensitivities. Recall that a finite-difference estimate of the sensitivity with respect to parameter  $c_k$ , over  $R$  coupled perturbed and unperturbed paths, is

$$\frac{\partial}{\partial c_k} \mathbb{E}[X(t, c)] \approx Z_R = \frac{1}{R} \sum_{r=1}^R \frac{X_{[r]}(t, c_k + \theta) - X_{[r]}(t, c_k)}{\theta}.$$

While we compute the coupled trajectories using the CFD, CRP, or CRN strategies, our method can be applied to other finite-difference sensitivity estimators [29].

The measure (12) can be calculated to rank parameters from most to least influential. Small values of  $\delta^{msqr}$  correspond to parameters with a small influence on the model. We select those parameters that show the value of  $\delta^{msqr}$  larger than 0.2 [39]. With an initial ranked list, we compute the collinearity indices for this list. This method can be applied to models of moderate size.

Algorithm 1 calculates the normalized sensitivity matrix, as follows. A grid with  $L$  time-points ranging from 0 to  $T$  is selected. We choose equally distributed time steps, such that data is collected from all important regions of the interval of integration. This depends on the particular model. We note that an adaptive time-stepping procedure can be included instead. Then, the sensitivity matrices  $S(t_l)$  from Equation (16) are approximated with a specific finite-difference sensitivity estimator. Afterwards, we compute the concatenated non-dimensional sensitivity matrix  $s$ . We normalize each column of  $s$  individually to ensure consistency and comparability. The normalization implies dividing each column  $s_k$  by its vectorial norm-2. Column normalization yields a matrix denoted by  $\tilde{S}$ . This matrix has as its  $k$ -th column  $\{\tilde{s}_k\} = s_k / \|s_k\|_2$ . Also, for each parameter  $c_k$  we compute the sensitivity measure  $\delta_k^{msqr}$  from Equation (12), using the entries of the  $k$ -th columns of the sensitivity matrices  $S(t_\ell)$ .

---

#### Algorithm 1 Computing the Normalized Sensitivity Matrix

---

**Initialize:** Time grid:  $0 = t_1 < t_2 < \dots < t_L = T$ .

**Input:** Estimates of sensitivity matrices  $S(t_\ell)$  from (16).

Compute the concatenated non-dimensional sensitivity matrix  $s$  from (18) with entries (17)

**for**  $k = 1$  to  $M$  **do**

normalize  $\tilde{s}_k = \frac{s_k}{\|s_k\|_2}$  where  $s_k$  is the  $k$ -th column of  $s$  and  $\|\cdot\|_2$  is norm-2

**end for**

Compute normalized matrix  $\tilde{S} = \{\tilde{s}_k\}_{1 \leq k \leq M}$

**for**  $k = 1$  to  $M$  **do**

Compute sensitivity measure  $\delta_k^{msqr}$  according to (12) for parameter  $c_k$

**end for**

---

In Algorithm 2, we introduce a method for the selection of identifiable parameter subsets based on sensitivity measures and collinearity indices. This procedure extends and refines a methodology by Gábor et al. [30] from the deterministic to the more difficult case of stochastic biochemical networks. The goal of Algorithm 2 is to iteratively assess the practical identifiability of subsets of model parameters. A threshold value is set for the collinearity indices, which measure the level of collinearity between parameter groups. The threshold value determines the acceptable level of collinearity. With a normalized sensitivity matrix obtained from Algorithm 1 as input, the following steps are considered. The parameters are ranked according to their sensitivity measure, those with a sensitivity measure below a critical value (chosen here as 0.2) are considered unimportant and may be discarded. If the ranked list of parameters is of moderate size, combinations of parameters are generated. For each combination, the algorithm computes the corresponding collinearity index. This involves calculating the collinearity indices for pairs, triples, etc. These indices quantify the degree of collinearity between the parameters of a certain group. When the computed collinearity index for a parameter subset is below the threshold value, that subset of parameters is deemed identifiable. By applying this algorithm, a subset of parameters with low collinearity and high identifiability can be selected. This allows for the reduction in model complexity and for the accurate and reliable estimation of the most important parameters, from the input data.

**Algorithm 2** Selecting a Subset of Identifiable Parameters

---

**Input:** Normalized sensitivity matrix;  
**Input:** Set threshold value of collinearity index:  $CI_{cr} = 20$   
**Require:** Rank parameters  $c_j$  based on  $\delta_j^{msqr} > 0.2$   
**if** Ranked list is of moderate size **then**  
    1: Number of all combinations:  $C = Length(combnk)$   
    2: Compute collinearity indices for all combinations of the ranked list of parameters:  
    **for**  $k = 1$  to  $C$  **do**  
        For every combination of the ranked list of parameters, calculate the collinearity indices:  
         $CI_2 = collinearityindex(pairs)$ ,  $CI_3 = collinearity(triples)$ , etc.  
         $L_2 = pair\ combination$ ,  $L_3 = triple\ combination$ , etc.  
    **end for**  
**end if**  
**if**  $CI_k \leq CI_{cr}$  **then**  
    The corresponding combination recorded as an identifiable set  
**end if**

---

**3. Results**

In this section, we apply our method to select subsets of practically identifiable parameters in the Chemical Master Equation on three realistic models. We observe that the collinearity indices play a significant role in finding the subsets of estimable parameters, using local stochastic sensitivities. The parametric sensitivities of the stochastic discrete model of well-stirred biochemical systems are approximated by finite-difference schemes, namely the Common Random Number, Common Reaction Path, and Coupled Finite Difference techniques. By applying perturbation in each of these finite-difference techniques, we can assess the sensitivity of the model outputs to changes in the model's parameters. The choice of perturbation size for finite-difference approximations is essential for obtaining accurate and reliable results while minimizing computational effort. The specific perturbation sizes, representing 5%, 1%, 2% of the parameter value, are often chosen based on a trade-off between accuracy and numerical stability. In addition, we find the parameters with high sensitivities. Those with low sensitivity have a reduced impact on the model outputs and cannot be estimated accurately. In the stochastic context, we consider the SVD of the normalized sensitivity matrix to determine its rank. This rank gives the number of model parameters that are not collinear.

For validation of the methods introduced above, we compare the results obtained with the Chemical Master Equation, with those derived with the Chemical Langevin Equation and those for the reaction rate equations, on two models of biochemically reacting systems. Still, we emphasize the importance of considering stochastic discrete models of biochemical networks to accurately describe the dynamics of these systems, particularly when some molecular populations are small or noise is driving the system behavior. The parametric sensitivities estimated for the reaction rate equations or the Chemical Langevin Equations may not yield accurate estimability results, in general. For each model, we generated 10,000 coupled trajectories to approximate the parametric sensitivities of the Chemical Master Equation by finite-difference schemes. The CFD strategy is considered to be more accurate and reliable than the CRN and the CRP methods [28]. The case studies tested are an infectious disease network [40], the Michaelis–Menten system and a genetic toggle-switch model [11].

**3.1. Infectious Disease Model**

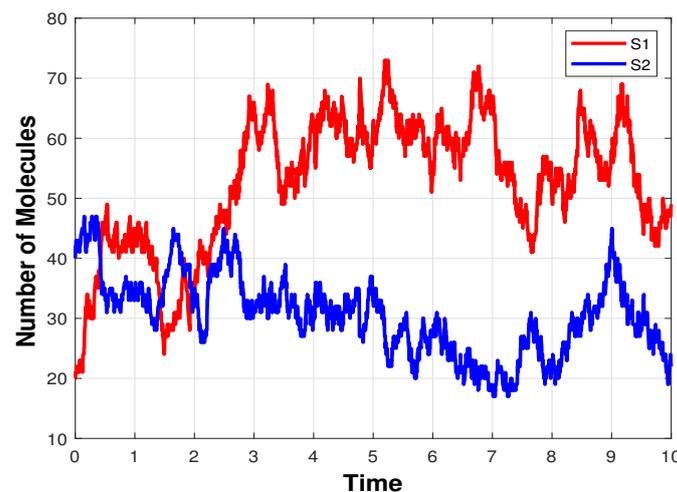
An infectious disease model [40] considers two species:  $S_1$ —the infected particles and  $S_2$ —the particles which can be infected. These species, which may depict molecules, cells, or humans, participate in five reactions. The first two reactions represent the death of species  $S_1$  and  $S_2$ , respectively, while the third and fourth reactions describe the birth

or production of particles of the  $S_1$  and  $S_2$  type. The two species interact through the fifth reaction, in which an infected particle  $S_1$  infects a particle  $S_2$ . The initial conditions are  $S_1(0) = 20$  and  $S_2(0) = 40$ . The system is studied on the time-interval  $[0, 10]$ . For our simulations, 10,000 trajectories were generated to estimate the solution of the Chemical Master Equation.

Table 1 provides information on the reaction channels of the biochemical system and the values of their rate parameters. It includes the reaction channels denoted by  $R_1$ ,  $R_2$ ,  $R_3$ ,  $R_4$ , and  $R_5$ . Each reaction is described by its reactants and products. The last column lists the parameter values corresponding to the rates at which the reactions occur. These parameter values are specified for the stochastic model considering molecular numbers, rather than for the deterministic reaction rate equations expressed in terms of concentrations. A sample trajectory of the number of the infected  $S_1$  particles and of the susceptible  $S_2$  particles as functions of time, computed using Gillespie's algorithm, is given in Figure 1.

**Table 1.** Infectious disease model: the list of reactions and the corresponding rate parameter values.

	Reaction Channel	Rate Parameter Value
$R_1$ :	$S_1 \rightarrow \emptyset$	$c_1 = 2.0$
$R_2$ :	$S_2 \rightarrow \emptyset$	$c_2 = 0.1$
$R_3$ :	$\emptyset \rightarrow S_1$	$c_3 = 25$
$R_4$ :	$\emptyset \rightarrow S_2$	$c_4 = 75$
$R_5$ :	$S_1 + S_2 \rightarrow S_1 + S_1$	$c_5 = 0.05$



**Figure 1.** Infectious disease model: the evolution in time of the number of molecules of the species  $S_1$ —infected individuals and  $S_2$ —individuals which can be infected, generated with Gillespie's algorithm, on the interval  $[0, 10]$ .

The finite-difference sensitivity estimations are calculated with 10,000 trajectories using the CFD, the CRN, and the CRP strategies, with a perturbation of 5% of the parameter value. The path-wise sensitivities for the Chemical Langevin Equation are computed over 10,000 trajectories, with the Euler-Maruyama scheme applied to the Equations (2) and (7), and are utilized to estimate the sensitivities of the expected value of the state vector. Also, the parametric sensitivities are approximated for the reaction rate equations. These estimations are used to calculate the collinearity indices for all parameter combinations, for the Chemical Master Equation, the Chemical Langevin Equation, and the RRE models. The results are presented in Tables 2–6. The sensitivity measures are reported in Table 2, showing that  $c_2$  is the least significant among all the parameters.

**Table 2.** Infectious disease model: comparison of  $\delta^{msqr}$  for a 5% perturbation.

Parameter	$\delta^{msqr}$ of CFD Sensitivity	$\delta^{msqr}$ of CRP Sensitivity	$\delta^{msqr}$ of CRN Sensitivity	$\delta^{msqr}$ of RRE Sensitivity	$\delta^{msqr}$ Path-Wise Sensitivity
$c_1$	0.97	0.96	0.94	0.97	0.98
$c_2$	0.02	0.02	0.1	0.02	0.02
$c_3$	0.26	0.29	0.26	0.26	0.26
$c_4$	0.55	0.66	0.54	0.55	0.55
$c_5$	0.68	0.69	0.67	0.71	0.71

Tables 3–6 reveal that the collinearity indices for the reaction rate equation and the Chemical Langevin Equation models exhibit greater consistency with the collinearity indices for the Chemical Master Equation, computed using with the CFD sensitivity estimator, compared to the CRN and the CRP estimators. Notably, the pair subset  $\{c_1, c_3\}$  has the highest collinearity index; however, it is relatively low for the CRP and the CRN schemes in comparison with the other estimations. This is due to the lower accuracy of the CRP and the CRN schemes when compared to the CFD technique. For pair sets, the subset  $\{c_1, c_3\}$ , for the triple sets, the subset  $\{c_3, c_4, c_5\}$  and among the quadruple ones, the subset  $\{c_2, c_3, c_4, c_5\}$  have high value of collinearity indices in relation to the other subsets.

**Table 3.** Infectious disease model: collinearity indices for pair subsets. The CME sensitivities are estimated over 10,000 trajectories with the CFD, CRN, and CRP algorithms and a 5% perturbation.

Parameter Subset	Collinearity Index of CFD Sensitivity	Collinearity Index of CRP Sensitivity	Collinearity Index of CRN Sensitivity	Collinearity Index of RRE Sensitivity	Collinearity Index of Path-Wise Sensitivity
$c_4 c_5$	1.18	1.13	1.18	1.2	1.19
$c_3 c_5$	1.93	1.17	1.94	1.92	1.95
$c_3 c_4$	1.339	1.25	1.32	1.32	1.31
$c_2 c_5$	1.103	1.15	1.13	1.18	1.17
$c_2 c_4$	4.69	2.37	1.16	9.77	9.96
$c_2 c_3$	1.43	1.27	1.02	1.34	1.33
$c_1 c_5$	1.86	1.89	1.9	1.85	1.86
$c_1 c_4$	1.35	1.28	1.33	1.34	1.34
$c_1 c_3$	10.816	3.04	7.2	11.34	11.22
$c_1 c_2$	1.466	1.31	1.00	1.36	1.35

There is no subset with high collinearity indices ( $>20$ ) in pair subsets (Table 3) but there is a parameter subset of size 3 with collinearity index greater than 20 (Table 4). In fact, the parameter subset  $\{c_3, c_4, c_5\}$  is not identifiable with the Coupled Finite Difference sensitivity estimator, the Chemical Langevin Equation, or the deterministic sensitivities. However, the Common Random Number and the Common Reaction Path sensitivities show different results. In Table 5, two parameter subsets of size 4 show a collinearity index greater than 20 with the deterministic, stochastic continuous, and CFD sensitivity estimations. All subsets containing the parameters  $\{c_3, c_4, c_5\}$  are collinear, which is in agreement with the results in Table 4. This indicates that these parameter subsets are poorly identifiable. Consequently, the sensitivity-based estimability analysis performed on the RRE, the CLE, and the CME models are in agreement, thus validating the proposed method for the more general discrete stochastic model. The Common Random Number and the Common Reaction Path techniques could not provide an accurate assessment of the identifiability of various subsets, with only 10,000 realizations, being thus less reliable.

**Table 4.** Infectious disease model: collinearity indices for triple subsets. The CME sensitivities are estimated over 10,000 trajectories with the CFD, CRN, and CRP algorithms and a 5% perturbation.

Parameter Subset	Collinearity Index of CFD Sensitivity	Collinearity Index of CRP Sensitivity	Collinearity Index of CRN Sensitivity	Collinearity Index of RRE Sensitivity	Collinearity Index of Path-Wise Sensitivity
$c_3 c_4 c_5$	21.19	2.63	9.6	21.3	21.77
$c_2 c_4 c_5$	5.0444	2.38	1.2	9.97	10.15
$c_2 c_3 c_5$	7.7768	2.91	2.01	10.48	10.51
$c_2 c_3 c_4$	4.88	2.38	1.43	9.83	10.01
$c_1 c_4 c_5$	9.92	3.65	9.4	10.83	10.98
$c_1 c_3 c_5$	11.07	3.12	7.2	11.68	11.73
$c_1 c_3 c_4$	10.87	3.05	7.2	11.46	11.45
$c_1 c_2 c_5$	7.44	4.8	2	7.87	7.95
$c_1 c_2 c_4$	4.95	2.38	1.43	9.82	10.01
$c_1 c_2 c_3$	11.02	3.06	7.3	11.45	11.44

**Table 5.** Infectious disease model: collinearity indices for quadruple subsets. The CME sensitivities are estimated over 10,000 trajectories with the CFD, CRN, and CRP algorithms and a 5% perturbation.

Parameter Subset	Collinearity Index of CFD Sensitivity	Collinearity Index of CRP Sensitivity	Collinearity Index of CRN Sensitivity	Collinearity Index of RRE Sensitivity	Collinearity Index of Path-Wise Sensitivity
$c_1 c_2 c_3 c_4$	11.509	3.06	7.3	11.53	11.49
$c_1 c_2 c_3 c_5$	11.092	4.88	7.3	13.65	13.53
$c_1 c_2 c_4 c_5$	10.2347	4.94	9.4	13.82	14.20
$c_1 c_3 c_4 c_5$	22.6313	3.87	10.54	22.19	22.49
$c_2 c_3 c_4 c_5$	21.4369	2.91	9.6	25.71	27.77

**Table 6.** Infectious disease model: collinearity indices for the set of all kinetic parameters. The CME sensitivities are estimated over 10,000 trajectories with the CFD, CRN, and CRP algorithms and a 5% perturbation. The singular values for the CFD, the CLE, and the RRE sensitivity estimations show that the number of parameters that are not collinear is four.

Parameter Subset	Collinearity Index of CFD Sensitivity	Collinearity Index of CRP Sensitivity	Collinearity Index of CRN Sensitivity	Collinearity Index of RRE Sensitivity	Collinearity Index of Path-Wise Sensitivity
$c_1 c_2 c_3 c_4 c_5$	22.65	5.01	10.54	26.17	28.09
singular values	16.31, 9.48, 1.06, 0.21, 0.06	16.27, 10.35, 2.98, 1.79, 0.14	15.86, 9.28, 1.31, 1.1, 0.52	36.73, 21.86, 2.21, 0.48, 0.09	37.03, 21.76, 2.19, 0.48, 0.09

### 3.2. Michaelis–Menten Model

The second model we analyze is the Michaelis–Menten biochemical system, which involves four species—a substrate  $S_1$ , an enzyme  $S_2$ , a complex  $S_3$  and a product  $S_4$ —and three reactions. We denote by  $Y_i$  the number of molecules of the species  $S_i$ . With this notation, the initial conditions for the number of molecules are  $Y_1(0) = [5 \times 10^{-7} n_A vol]$ ,  $Y_2(0) = [2 \times 10^{-7} n_A vol]$  and  $Y_3(0) = Y_4(0) = 0$ , where  $n_A = 6.023 \times 10^{23}$  is Avogadro's number and  $vol = 10^{-15}$  denotes the volume of the system. The reactions and the values of the rate parameters are included in Table 7. This model is integrated on the interval  $[0, 50]$ . Figure 2 depicts a realization of the system state, simulated with Gillespie's algorithm.

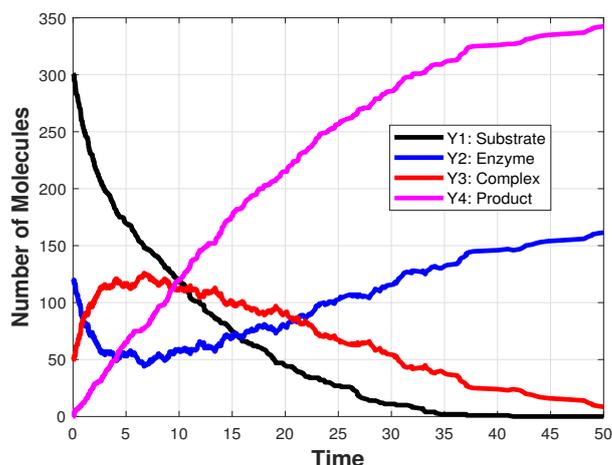


Figure 2. Michaelis–Menten model: the evolution in time of the number of molecules of a substrate, an enzyme, a complex and a product, generated with Gillespie’s algorithm, on the interval [0, 50].

Table 7. Michaelis–Menten model: the list of reactions and the corresponding rate parameter values.

	Reaction Channel	Rate Parameter Value
$R_1:$	$S_1 + S_2 \rightarrow S_3$	$c_1 = 10^6 / (n_A vol)$
$R_2:$	$S_3 \rightarrow S_1 + S_2$	$c_2 = 10^{-4}$
$R_3:$	$S_3 \rightarrow S_4 + S_2$	$c_3 = 10^{-1}$

We start by approximating the parametric sensitivities for the Chemical Master Equation. The finite-difference sensitivity estimations obtained with the CFD, the CRP, and the CRN algorithms use a perturbation which represents 1% and 5%, respectively, of the value of the parameter of interest. The sensitivity measures provided in Table 8 indicate that  $c_2$  may not be estimated as accurately as the other parameters. The collinearity indices obtained for the perturbation value 1% with each sensitivity estimator for pairs of parameters are reported in Table 9, while the indices for the set of all parameters are recorded in Table 10. For each subset, the results for the stochastic Michaelis–Menten model demonstrate low collinearity indices, below 20. The choice of the finite-difference sensitivity estimator does not significantly affect the parameter identifiability. The stochastic discrete modeling approach to identifiability analysis yields parameter subsets that are not collinear for the Michaelis–Menten system. Additionally, the Tables include the RRE identifiability metrics to validate the CME estimability results. The collinearity indices for the perturbation value of 5% can be found in the Appendix A, and they are consistent with the results obtained using a perturbation of 1%.

Table 8. Michaelis–Menten model: comparison of  $\delta^{msqr}$  for a 1% perturbation.

Parameter	$\delta^{msqr}$ of CFD Sensitivity	$\delta^{msqr}$ of CRP Sensitivity	$\delta^{msqr}$ of CRN Sensitivity	$\delta^{msqr}$ of RRE Sensitivity
$c_1$	1.11	1.1	1.07	1.07
$c_2$	0.002	0.01	0.003	0.002
$c_3$	1.31	1.30	1.29	1.29

**Table 9.** Michaelis–Menten model: collinearity indices for pair subsets. The CME sensitivities are estimated over 10,000 trajectories with the CFD, CRN, and CRP algorithms and a 1% perturbation.

Parameter Subset	Collinearity Index of CFD Sensitivity	Collinearity Index of CRP Sensitivity	Collinearity Index of CRN Sensitivity	Collinearity Index of RRE Sensitivity
$c_1 c_2$	2.9	1.35	1.47	4.85
$c_1 c_3$	2.21	2.17	2.17	2.17
$c_2 c_3$	1.56	1.21	1.2	1.87

**Table 10.** Michaelis–Menten model: collinearity indices for the triple subset. The CME sensitivities are estimated over 10,000 trajectories with the CFD, CRN, and CRP algorithms and a 1% perturbation.

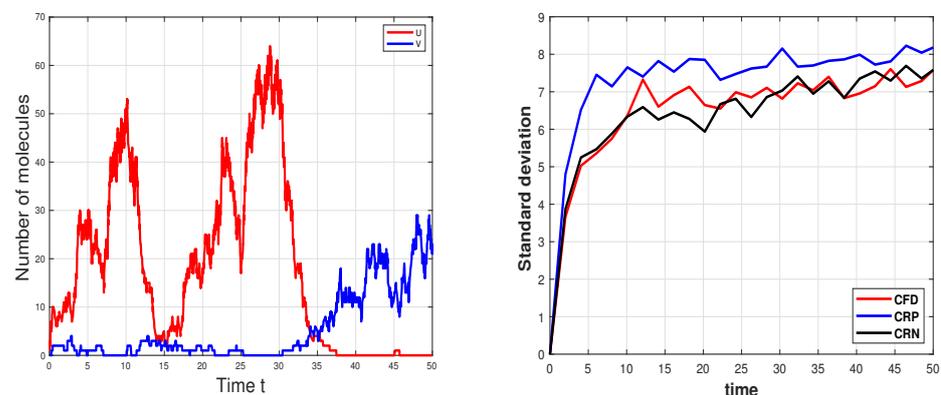
Parameter Subset	Collinearity Index of CFD Sensitivity	Collinearity Index of CRP Sensitivity	Collinearity Index of CRN Sensitivity	Collinearity Index of RRE Sensitivity
$c_1 c_2 c_3$	3.92	2.25	2.43	5.3

### 3.3. Genetic Toggle Switch Model

The last biochemical system investigated is the genetic toggle switch [11,28]. Multi-stable stochastic switches arise in modeling key biological processes. The model considers two gene pairs, whose interaction creates a bistable switch, as each gene negatively regulates the synthesis of the other gene. Due to the presence of noise, the system can transition between the states represented by an abundance of one species and an almost total absence of the other. In this genetic switch system, the two species  $U$  and  $V$  take part in four reactions. Table 11 specifies the reaction channels and their propensities. We examine the system using the following parameter values [11]

$$\alpha_1 = 50, \beta = 2.5, \alpha_2 = 16, \gamma = 1, \quad (20)$$

and the initial conditions  $X_V(0) = X_U(0) = 0$ . Figure 3 displays a sample path for the molecular numbers of the two species, simulated with Gillespie’s algorithm (left) along with the standard deviation of the CFD, CRP, and CRN sensitivity estimators as functions of time (right).



**Figure 3.** Genetic toggle switch model: (Left): the evolution in time of the number of molecules of the species  $U$  and  $V$ , generated with Gillespie’s algorithm, on the interval  $[0, 50]$ . (Right): standard deviations of the three estimators, CFD, CRP, and CRN.

**Table 11.** Genetic toggle switch model: the list of reactions and their propensity functions.

	Reaction Channel	Propensity Function
$R_1:$	$\emptyset \rightarrow U$	$a_1 = \frac{\alpha_1}{1 + X_V^\beta}$
$R_2:$	$U \rightarrow \emptyset$	$a_2 = X_U$
$R_3:$	$\emptyset \rightarrow V$	$a_3 = \frac{\alpha_2}{1 + X_U^\gamma}$
$R_4:$	$V \rightarrow \emptyset$	$a_4 = X_V$

The reaction rate equation model cannot capture the stochastic transitions between the states, and thus the deterministic tools for analyzing this system are not applicable. We perform an estimability analysis of the Chemical Master Equation model for the genetic toggle switch, on the interval  $[0, 50]$ . To assess how variations in the parameter values affect the dynamics of the system, we approximate the local sensitivities with respect to the parameters whose values are given by (20). We simulate 10,000 coupled sample paths with the CFD, and the CRP methods. The finite-difference sensitivity estimators are applied with a perturbation  $\theta = 10^{-4}$  for each parameter value. The sensitivity measures are provided in Table 12 and those calculated using the CFD method show that all parameters have  $\delta^{msqr} > 0.2$ , being thus important enough, while the RRE sensitivity measures indicate that the parameters  $\beta$  and  $\gamma$  are insignificant.

**Table 12.** Genetic toggle switch model: comparison of  $\delta^{msqr}$ .

Parameter	$\delta^{msqr}$ of CFD Sensitivity	$\delta^{msqr}$ of RRE Sensitivity
$\alpha_1$	2.22	0.89
$\beta$	0.6762	0
$\alpha_2$	4.21	0.31
$\gamma$	4.3	0

Employing the local sensitivity approximations, we compute the collinearity indices for all the subsets of the parameter set  $\{\alpha_1, \alpha_2, \beta, \gamma\}$ . Tables 13–15 record the collinearity indices for the pair, triple and quadruple subsets, respectively. No subset of parameters exhibits collinearity based on the CFD, the CRP, and the CRN sensitivity estimations. We conclude that all four parameters are identifiable for the stochastic discrete model. These results are confirmed by the singular values computed with the CFD sensitivity estimator, which are  $[32.21; 29; 12.18; 4]$ . Different values of the parameters for this model may yield different results for estimability in the stochastic genetic toggle-switch system.

**Table 13.** Genetic toggle switch model: collinearity indices for pair subsets.

Parameter Subset	Collinearity Index of CFD Sensitivity	Collinearity Index of CRP Sensitivity	Collinearity Index of CRN Sensitivity	Collinearity Index of RRE Sensitivity
$\alpha_1 \alpha_2$	1	1.01	1.72	2.22
$\gamma \alpha_2$	1.32	1.08	1.12	*
$\gamma \alpha_1$	1.27	1.17	1.07	*
$\beta \alpha_2$	1.01	1.1	1.56	*
$\beta \alpha_1$	1.00	1.35	2.13	*
$\beta \gamma$	1.19	1.25	1.1	*

The CME sensitivities with respect to parameters are estimated over 10,000 with the CFD and CRP methods and perturbation  $\theta = 10^{-4}$ . \*: Collinearity index does not exist.

**Table 14.** Genetic toggle switch model: collinearity indices for triple subsets.

Parameter Subset	Collinearity Index of CFD Sensitivity	Collinearity Index of CRP Sensitivity	Collinearity Index of CRN Sensitivity	Collinearity Index of RRE Sensitivity
$\beta \gamma \alpha_1$	1.38	1.37	2.46	*
$\beta \gamma \alpha_2$	1.42	1.25	1.80	*
$\beta \alpha_1 \alpha_2$	1.01	1.38	2.18	*
$\gamma \alpha_1 \alpha_2$	1.52	1.19	1.73	*

The CME sensitivities with respect to parameters are estimated over 10,000 with the CFD and CRP methods and perturbation  $\theta = 10^{-4}$ . \*: Collinearity index does not exist.

**Table 15.** Genetic toggle switch model: collinearity indices for the quadruple subset.

Parameter Subset	Collinearity Index of CFD Sensitivity	Collinearity Index of CRP Sensitivity	Collinearity Index of CRN Sensitivity	Collinearity Index of RRE Sensitivity
$\beta \gamma \alpha_1 \alpha_2$	1.64	1.39	2.45	*

The CME sensitivities with respect to parameters are estimated over 10,000 with the CFD and CRP methods and perturbation  $\theta = 10^{-4}$ . \*: Collinearity index does not exist.

#### 4. Discussion

Stochastic models of well-stirred biochemical processes provide a valuable framework for capturing inherent variability at the cellular level when some molecular species have low amounts. Chemical Master Equation is a frequently adopted stochastic discrete model for such processes. By contrast, deterministic approaches are often not suitable for modeling cellular systems as they fail to capture the intrinsic randomness observed experimentally. Many models of realistic biochemical processes depend on a fairly large number of parameters. The values of some of these parameters may be unknown and have to be estimated. Parameter estimation is a critical step in modeling biochemical systems. However, determining appropriate parameter values for stochastic discrete models of biochemical networks poses many challenges. It is essential to determine the key parameters which are identifiable from the experimental data, as well as those that cannot be reliably estimated. For a subset of parameters to be practically identifiable, each parameter of the subset should have a significant contribution to the system dynamics as well as the parameters of the subset should not be correlated.

In this work, we propose a method for detecting collinearity in subsets of parameters for the stochastic discrete model of the Chemical Master Equation, with the goal of finding the parameter sets that exert the greatest influence on the biochemical system state. In addition, we introduce a technique for determining the highest parameter identifiable sets for stochastic biochemical systems, by extending methods from deterministic models to stochastic models. Our analysis is based on estimating the local sensitivities of the system state with respect to the model's parameters. This is achieved by utilizing finite-difference approximations of the parameter sensitivities, specifically the Coupled Finite Difference, the Common Reaction Path, and the Common Random Number schemes. Furthermore, we examine the role of the singular value decomposition of the sensitivity matrix in identifying parameters that are not collinear in stochastic models of biochemical systems. On one hand, we showed that our practical identifiability method is accurate, by comparing the results obtained in the deterministic and stochastic scenarios, on two biochemical systems of practical importance, for which the deterministic model accurately describes the evolution of the expected value of the stochastic system state. Excellent agreement among the various approaches was obtained for these biochemical networks. On the other hand, we wish to emphasize that, in general, a stochastic strategy for selecting identifiable parameter sets should be considered, as it relies on more accurate and reliable estimations of the parametric sensitivities for the widely applicable model of the Chemical Master Equation,

compared to the deterministic reaction rate equations. The advantages of our approach over the deterministic one were illustrated by the tests performed on a third model, a genetic toggle switch system exhibiting an interesting multistable behavior. For this model, our stochastic identifiability strategies display excellent performance, while the deterministic techniques show their limitations, by not being able to assess the estimability of the model parameters.

We expect the method to perform best on stochastic biochemical models with a moderate number of reaction rate parameters. Specifying identifiable parameter subsets with the tools provided above may be used to refine models, improve predictions, and study the underlying biological processes under consideration.

**Author Contributions:** Conceptualization, S.G. and S.I.; methodology, S.G. and S.I.; software, S.G.; validation, S.I.; investigation, S.G. and S.I.; writing—original draft preparation, S.G. and S.I.; writing—review and editing, S.I.; visualization, S.G. and S.I.; supervision, S.I.; funding acquisition, S.I. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by a grant from the National Sciences and Engineering Research Council of Canada (NSERC)—Grant No. RGPIN-2020-05469, and Toronto Metropolitan University.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** No new data were created.

**Acknowledgments:** The authors wish to thank the anonymous referees for their suggestions which helped improve the presentation and Monjur Morshed for discussions.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

**Table A1.** Michaelis–Menten model: comparison of  $\delta^{msqr}$  for a 5% perturbation.

Parameter	$\delta^{msqr}$ of CFD Sensitivity	$\delta^{msqr}$ of CRP Sensitivity	$\delta^{msqr}$ of CRN Sensitivity	$\delta^{msqr}$ of RRE Sensitivity
$c_1$	1.03	1.04	1.04	1.07
$c_2$	0.002	0.005	0.003	0.002
$c_3$	1.22	1.23	1.23	1.29

**Table A2.** Michaelis–Menten model: collinearity indices for pair subsets. The CME sensitivities are estimated over 10,000 trajectories with the CFD, CRN and CRP algorithms and a 5% perturbation.

Parameter Subset	Collinearity Index of CFD Sensitivity	Collinearity Index of CRP Sensitivity	Collinearity Index of CRN Sensitivity	Collinearity Index of RRE Sensitivity
$c_1 c_2$	3.43	2.18	2.59	4.85
$c_1 c_3$	2.21	2.13	2.13	2.17
$c_2 c_3$	1.67	1.48	1.49	1.87

**Table A3.** Michaelis–Menten model: collinearity indices for the triple subset. The CME sensitivities are estimated over 10,000 trajectories with the CFD, CRN and CRP algorithms and a 5% perturbation.

Parameter Subset	Collinearity Index of CFD Sensitivity	Collinearity Index of CRP Sensitivity	Collinearity Index of CRN Sensitivity	Collinearity Index of RRE Sensitivity
$c_1 c_2 c_3$	4.08	2.78	3.4	5.3

## References

1. Kitano, H. Computational systems biology. *Nature* **2002**, *420*, 206–210. [[CrossRef](#)] [[PubMed](#)]
2. Maheshri, N.; O’Shea, E.K. Living with noisy genes: How cells function reliably with inherent variability in gene expression. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 413–434. [[CrossRef](#)]
3. Raj, A.; van Oudenaarden, A. Nature, nurture, or chance: Stochastic gene expression and its consequences. *Cell* **2008**, *135*, 216–226. [[CrossRef](#)] [[PubMed](#)]
4. Ozbudak, E.M.; Thattai, M.; Kurtser, I.; Grossman, A.D.; van Oudenaarden, A. Regulation of noise in the expression of a single gene. *Nat. Genet.* **2002**, *31*, 69–73. [[CrossRef](#)]
5. Raser, J.M.; O’Shea, E.K. Noise in gene expression: Origins, consequences, and control. *Science* **2005**, *309*, 2010–2013. [[CrossRef](#)]
6. Gillespie, D.T. A rigorous derivation of the chemical master equation. *Stat. Mech. Its Appl.* **1992**, *188*, 404–425. [[CrossRef](#)]
7. McQuarrie, D.A. Stochastic approach to chemical kinetics. *J. Appl. Probab.* **1967**, *4*, 413. [[CrossRef](#)]
8. Gillespie, D.T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* **1976**, *22*, 403–434. [[CrossRef](#)]
9. Gillespie, D.T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **1977**, *81*, 2340–2361. [[CrossRef](#)]
10. Ethier, S.N.; Kurtz, T.G. *Markov Processes: Characterization and Convergence*; Wiley: New York, NY, USA, 1986.
11. Rathinam, M.; Sheppard, P.W.; Khammash, M. Efficient computation of parameter sensitivities of discrete stochastic chemical reaction networks. *J. Chem. Phys.* **2010**, *132*, 034103–034116. [[CrossRef](#)]
12. El Samad, H.; Khammash, M.; Petzold, L.; Gillespie, D.T. Stochastic modelling of gene regulatory networks. *Int. J. Robust Nonlinear Control* **2005**, *15*, 691–711. [[CrossRef](#)]
13. Strehl, R.; Ilie, S. Hybrid stochastic simulation of reaction-diffusion systems with slow and fast dynamics. *J. Chem. Phys.* **2015**, *143*, 234108. [[CrossRef](#)] [[PubMed](#)]
14. Wilkinson, D.J. *Stochastic Modelling for Systems Biology*; Taylor & Francis: Boca Raton, FL, USA, 2019.
15. Thanh, V.H.; Zunino, R.; Priami, C. On the rejection-based algorithm for simulation and analysis of large-scale reaction networks. *J. Chem. Phys.* **2015**, *142*, 244106. [[CrossRef](#)]
16. Barrows, D.; Ilie, S. Parameter estimation for the reaction-diffusion master equation. *AIP Adv.* **2023**, *13*, 065318. [[CrossRef](#)]
17. Petre, I.; Mizera, A.; Hyder, C.L.; Meinander, A.; Mikhailov, A.; Morimoto, R.I.; Sistonen, L.; Eriksson, J.E.; Back, R.J. A simple mass-action model for the eukaryotic heat shock response and its mathematical validation. *Nat. Comput.* **2011**, *10*, 595–612. [[CrossRef](#)]
18. Vajda, S.; Rabitz, H.; Walter, E.; Lecourtier, Y. Qualitative and quantitative identifiability analysis of nonlinear chemical kinetic models. *Chem. Eng. Commun.* **1989**, *83*, 191–219. [[CrossRef](#)]
19. Brun, R.; Reichert, P.; Künsch, H.R. Practical identifiability analysis of large environmental simulation models. *Water Resour. Res.* **2001**, *37*, 1015–1030. [[CrossRef](#)]
20. Brun, R.; Kühni, M.; Siegrist, H.; Gujer, W.; Reichert, P. Practical identifiability of ASM2d parameters—Systematic selection and tuning of parameter subsets. *Water Res.* **2002**, *36*, 4113–4127. [[CrossRef](#)] [[PubMed](#)]
21. Chis, O.T.; Banga, J.R.; Balsa-Canto, E. Structural identifiability of systems biology models: A critical comparison of methods. *PLoS ONE* **2011**, *6*, e27755. [[CrossRef](#)] [[PubMed](#)]
22. Holmberg, A. On the practical identifiability of microbial growth models incorporating Michaelis-Menten type nonlinearities. *Math. Biosci.* **1982**, *62*, 23–43. [[CrossRef](#)]
23. Jacquéz, J.; Greif, P. Numerical parameter identifiability and estimability: Integrating identifiability, estimability, and optimal sampling design. *Math. Biosci.* **1985**, *77*, 201–227. [[CrossRef](#)]
24. Komorowski, M.; Costa, M.J.; Rand, D.A.; Stumpf, M.P.H. Sensitivity, robustness, and identifiability in stochastic chemical kinetics models. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 8645–8650. [[CrossRef](#)] [[PubMed](#)]
25. Rodriguez-Fernandez, M.; Egea, J.A.; Banga, J.R. Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems. *BMC Bioinform.* **2006**, *7*, 483. [[CrossRef](#)]
26. Villaverde, A.F. Observability and structural identifiability of nonlinear biological systems. *Complexity* **2019**, *2019*, 1–12. [[CrossRef](#)]
27. Anderson, D.F. An efficient finite-difference method for parameter sensitivities of continuous time Markov chains. *SIAM J. Num. Anal.* **2012**, *50*, 2237–2258. [[CrossRef](#)]
28. Srivastava, R.; Anderson, D.F.; Rawlings, J.B. Comparison of finite difference based methods to obtain sensitivities of stochastic chemical kinetic models. *J. Chem. Phys.* **2013**, *138*, 074110. [[CrossRef](#)] [[PubMed](#)]
29. Morshed, M. Efficient Finite-Difference Methods for Sensitivity Analysis of Stiff Stochastic Discrete Models of Biochemical Systems. Ph.D. Thesis, University of Waterloo, Waterloo, ON, Canada, 2017.
30. Gábor, A.; Villaverde, A.F.; Banga, J.R. Parameter identifiability analysis and visualization in large-scale kinetic models of biosystems. *BMC Syst. Biol.* **2017**, *11*, 54. [[CrossRef](#)] [[PubMed](#)]
31. Turner, T.E.; Schnell, S.; Burrage, K. Stochastic approaches for modelling in vivo reactions. *Comput. Biol. Chem.* **2004**, *28*, 165–178. [[CrossRef](#)]
32. Gillespie, D.T. The chemical Langevin equations. *J. Phys. Chem.* **2000**, *113*, 297–306. [[CrossRef](#)]
33. Ilie, S.; Gholami, S. Simplifying stochastic mathematical models of biochemical systems. *Appl. Math.* **2013**, *4*, 248–256. [[CrossRef](#)]
34. Glasserman, P. *Monte Carlo Methods in Financial Engineering*; Springer: Berlin/Heidelberg, Germany, 2004.

35. Ilie, S. Variable time-stepping in the pathwise numerical solution of the chemical Langevin equation. *J. Phys. Chem.* **2012**, *137*, 234110. [[CrossRef](#)] [[PubMed](#)]
36. Sotiropoulos, V.; Kaznessis, Y.N. An adaptive time step scheme for a system of stochastic differential equations with multiple multiplicative noise: Chemical Langevin equation, a proof of concept. *J. Chem. Phys.* **2008**, *128*, 014103. [[CrossRef](#)] [[PubMed](#)]
37. Corless, R.M.; Fillion, N. *An Introduction to Numerical Methods from the Point of View of Backward Error Analysis*; Springer: New York, NY, USA, 2013.
38. Golub, G.; Van Loan, C. *Matrix Computations*, 3rd ed.; The Johns Hopkins University Press: London, UK, 1996.
39. Weijers, S.R.; Vanrolleghem, P.A. A procedure for selecting best identifiable parameters in calibrating activated sludge model no. 1 to full-scale plant data. *Water Sci. Technol.* **1997**, *36*, 69–79. [[CrossRef](#)]
40. Jahnke, T. On reduced models for chemical master equation. *Multiscale Model. Simul.* **2011**, *9*, 1646–1676. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.