

MULTISCALE GLOBAL SENSITIVITY ANALYSIS FOR STOCHASTIC CHEMICAL SYSTEMS*

MICHAEL MERRITT[†], ALEN ALEXANDERIAN[†], AND PIERRE A. GREMAUD[†]

Abstract. Sensitivity analysis is routinely performed on simplified surrogate models as the cost of such analysis on the original model may be prohibitive. Little is known in general about the induced bias on the sensitivity results. Within the framework of chemical kinetics, we provide a full justification of the above approach in the case of variance based methods provided the surrogate model results from the original one through the thermodynamic limit. In particular, we analyze convergence of the Sobol’ indices for observables in a stochastic chemical system to the corresponding Sobol’ indices from the system in the thermodynamic limit. We provide illustrative numerical examples in the context of a Michaelis–Menten system and a biochemical reaction network describing a genetic oscillator.

Key words. chemical reaction networks, stochastic processes, global sensitivity analysis, multiscale modeling, thermodynamic limit

AMS subject classifications. 65C20, 65Z05, 92E20, 80A30

DOI. 10.1137/20M1323989

1. Introduction. Striking a balance between accuracy and cost is one of the core challenges of scientific computing. A high fidelity, high cost model g is thus often replaced in practice by a lower cost model $\tilde{g}$, of (usually) lower fidelity, to enable the analysis of the application under study. The techniques to develop and construct surrogate models are many and range from approximation theory to physics [14]. The analysis of the original model g is then replaced by the analysis of a surrogate $\tilde{g}$ with the implicit assumption that

$$(1.1) \quad \text{if } g \approx \tilde{g}, \text{ then } \mathcal{I}(g) \approx \mathcal{I}(\tilde{g}),$$

where $\mathcal{I}$ represents some operation on g . The extent to which (1.1) is satisfied clearly depends on $\mathcal{I}$ and on the relationship between g and $\tilde{g}$. This paper is a first step toward the justification of (1.1) when $\mathcal{I}$ stands for the sensitivity of the model to its input parameters. We restrict our attention to an important family of physically based surrogates corresponding to $\tilde{g}$ being the thermodynamic limit of g and take chemical reaction networks as a motivating application. Recent results about approximation based—rather than physically based—surrogates can be found in [18].

Consider thus the evolution of a system of chemically reacting molecules; molecular dynamics simulation is the most faithful way of modeling such a system. There, each individual molecule and corresponding species population are tracked and chemical reactions are modeled as distinct events. Due to quantum effects and since such systems are typically not isolated, molecular populations are integer variables which evolve stochastically [8]. In spite of this, chemical kinetics is often analyzed using real—as opposed to integer—variables which evolve deterministically; that this is the

*Received by the editors March 9, 2020; accepted for publication (in revised form) November 16, 2020; published electronically March 11, 2021.

<https://doi.org/10.1137/20M1323989>

Funding: This work was supported by the National Science Foundation under grants DMS-1745654 and DMS-1953271.

[†]Department of Mathematics, North Carolina State University, Raleigh, NC USA 27695-8205 (mbmerritt@ncsu.edu, alexanderian@ncsu.edu, gremaud@ncsu.edu).

case is a testimony to the appeal of simplified low-cost models. *Stochastic* chemical kinetics is however necessary to the study of many cellular systems in biology where the relatively small molecular populations may preclude the use of simplified models obtained through the thermodynamic limit, i.e., in the limit of large volumes, and may require a stochastic rather than deterministic model.

Assume we have both a high-cost stochastic model g and a low-cost deterministic surrogate $\tilde{g}$ such that

$$(1.2) \quad q = g(\mathbf{k}, \omega), \quad \tilde{q} = \tilde{g}(\mathbf{k}), \quad \text{and } q \approx \tilde{q} \text{ in some sense,}$$

where the outcome ω corresponds to the intrinsic stochasticity of the model g , and q and $\tilde{q}$ are the respective quantities of interest (QoIs); here $\mathbf{k} = (k_1, \dots, k_M)$ is a list of shared uncertain parameters. As shown below, the field of chemical kinetics falls under this framework. A fundamental assumption we make in the present work is that the intrinsic model stochasticity is independent of the randomness in the uncertain parameters. This assumption, which also appears in related works [16, 17], is a natural one from the point of view of modeling under uncertainty: in the present setting, parametric uncertainty is due to lack of knowledge, i.e., it is epistemic, whereas model stochasticity is inherent to the system and aleatoric in nature.

Global sensitivity analysis (GSA) aims to quantify the relative importance of uncertain model parameters in determining the QoI [12, 13, 22, 26]. We analyze whether GSA can be performed on the surrogate $\tilde{g}$ rather than g and still yield information on the original model g . In other words, we are asking when the diagram in Figure 1.1 is commutative.

$$\begin{array}{ccc} q = g(\mathbf{k}, \omega) & \xrightarrow{\text{GSA}} & \{\mathcal{I}_j(\omega)\}_{j=1}^M \\ \text{limiting process} \downarrow & & \downarrow \text{limiting process} \\ \tilde{q} = \tilde{g}(\mathbf{k}) & \xrightarrow{\text{GSA}} & \{\tilde{\mathcal{I}}_j\}_{j=1}^M \end{array}$$

FIG. 1.1. Schematic representation of the question considered in this paper: For what type of limiting process is the diagram commutative? The model g is expensive to evaluate and stochastic, while the surrogate model $\tilde{g}$ is deterministic and cheap. We show that the diagram is commutative if the limiting process is the thermodynamic limit.

In Figure 1.1, $\mathcal{I}$ and $\tilde{\mathcal{I}}$ refer to importance indices from some GSA method; presumably, when applied to stochastic models, the GSA approach yields indices which themselves are random variables. This is, for instance, the case for variance based methods and Sobol' indices, which we use in this paper; see [10] and section 4. For chemical kinetics, the *limiting process* in the above diagram is the thermodynamic limit; see section 2. The above diagram does not in general commute; see [10] for simple analytical examples of noncommutativity when the limiting process linking the stochastic model to its surrogate is the expectation or some other ω -moment.

2. Chemical kinetics models. We consider chemical systems with N reacting species. We let $\mathbf{X}(t)$ be the state vector of a chemical system, where $X_i(t)$, the i th component of $\mathbf{X}(t)$, corresponds to the number of molecules of the i th species, $i = 1, \dots, N$, at time t .

2.1. The RTC representation. To guide our discussion, consider the simple case of one reaction and three species S_1 , S_2 , and S_3 ,



where one molecule of S_1 and one molecule of S_2 combine to produce one molecule of S_3 . The evolution of the state $\mathbf{X}(t) = [X_1(t) \ X_2(t) \ X_3(t)]^\top$ takes the form

$$(2.2) \quad \mathbf{X}(t) = \mathbf{X}(0) + \boldsymbol{\nu}R(t),$$

where $\boldsymbol{\nu} = [-1 \ -1 \ 1]^\top$ is the stoichiometric vector of that reaction (S_1 and S_2 lose one molecule and S_3 gains one) while $R(t)$ is the number of times the reaction takes place between time 0 and t . It is intuitive, and has been justified on physical grounds [8, 11], that the probability of the reaction occurring between time t and $t + dt$ is proportional to $X_1(t)$, $X_2(t)$, and dt , which suggests the model [2, 6]

$$(2.3) \quad R(t) = Y \left(\int_0^t c X_1(s) X_2(s) ds \right),$$

where c is a proportionality constant with dimension 1/time and Y is a unit-rate Poisson process: $Y(0) = 0$, Y has independent increments, and $Y(t+s) - Y(s)$ has a Poisson distribution with parameter t for all $t, s \geq 0$, i.e., $\mathbb{P}(Y(t+s) - Y(s) = n) = e^{-t} t^n / n!$.

More generally, the evolution of a system with N species and M reactions is governed by the propensity functions a_j , $j = 1, \dots, M$, where $a_j(\mathbf{X}(t)) dt$ represents the probability that the j th reaction occurs during the time interval $[t, t + dt)$. For instance, in the case of (2.1), the propensity function is $a(\mathbf{X}(t)) = c X_1(t) X_2(t)$. The resulting evolution equation, often referred to as the *random time change* (RTC) representation [1, 2, 4, 6], is then

$$(2.4) \quad \mathbf{X}(t) = \mathbf{X}(0) + \sum_{j=1}^M \boldsymbol{\nu}_j Y_j(\tau_j(t)),$$

$$(2.5) \quad \tau_j(t) = \int_0^t a_j(\mathbf{X}(s)) ds, \quad j = 1, \dots, M,$$

where $\boldsymbol{\nu}_j$ is the stoichiometric vector of the j th reaction and the Y_j 's are independent unit-rate Poisson processes. We follow [7] and define in (2.5) an internal time τ_j for each reaction; as the a_j 's have dimension 1/time, the τ_j 's are dimensionless.

The Law of Mass Action [2] leads to the propensity functions for the three main types of reactions:

$$(2.6) \quad S_m \rightarrow \text{something} \Rightarrow a_j(\mathbf{X}(t)) = c_j X_m(t),$$

$$(2.7) \quad S_m + S_n \rightarrow \text{something} \Rightarrow a_j(\mathbf{X}(t)) = c_j X_m(t) X_n(t) \quad \text{if } m \neq n,$$

$$(2.8) \quad S_m + S_m \rightarrow \text{something} \Rightarrow a_j(\mathbf{X}(t)) = c_j \frac{1}{2} X_m(t) (X_m(t) - 1).$$

The reactions (2.6), (2.7), and (2.8) are known as first order, second order, and dimerization reactions, respectively. The form of the propensity functions for other common reaction types can be found, for example, in [9].

2.2. The thermodynamic limit. In our analysis, we consider the limiting behavior of chemical systems as the system size approaches infinity. For example, as the system size increases, the likelihood of a particular reaction to fire may change in the event that certain molecules must interact. To this end, we aim to update the propensity functions by introducing a system size parameter V given by the product of the system volume and the Avogadro number n_A . As is common in the study of chemical systems, we write the stoichiometric vectors as follows:

$$\boldsymbol{\nu}_j = \boldsymbol{\nu}'_j - \boldsymbol{\nu}''_j, \quad j = 1, \dots, M,$$

where the entries of $\boldsymbol{\nu}'_j$ and $\boldsymbol{\nu}''_j$ are the number of molecules of system species that are created and consumed in the j th reaction, respectively. Following the notation of [29], we define the V -dependent propensity functions as follows:

$$a_j^V(\mathbf{x}) = \frac{k_j}{V^{\|\boldsymbol{\nu}''_j\| - 1}} \prod_{i=1}^N \binom{x_i}{\nu''_{ij}}, \quad j = 1, \dots, M,$$

where the k_j 's are reaction rate constants. The V -dependent system trajectory is described by the RTC representation,

$$(2.9) \quad \mathbf{X}^V(t) = V\mathbf{x}_0 + \sum_{j=1}^M \boldsymbol{\nu}_j Y_j \left(\int_0^t a_j^V(\mathbf{X}^V(s)) ds \right).$$

Here we have let $\mathbf{X}^V(0) = V\mathbf{x}_0$, where $\mathbf{x}_0 \in \mathbb{R}_{\geq 0}^N$ is a fixed vector. Throughout we will work with a sequence of V values such that $V\mathbf{x}_0$ is in $\mathbb{Z}_{\geq 0}^N$. Ensuring existence of such a sequence requires some assumptions on $\mathbf{x}_0$ and the nominal (initial) system volume. Specifically, in our study of limiting behavior of systems, we may assume that the system's nominal volume $\mathcal{V}_{\text{nom}}$ and $\mathbf{x}_0$ are such that $V_{\text{nom}}\mathbf{x}_0 = \mathcal{V}_{\text{nom}}n_A\mathbf{x}_0$ is a vector in $\mathbb{Z}_{\geq 0}^N$. We then consider a sequence of system sizes given by $V_m = mV_{\text{nom}}$, $m = 1, 2, \dots$.

Notice that the RTC formulation (2.9) is a restatement of (2.4), except with the dependence on system size made precise. For instance, considering the system at its nominal volume $\mathcal{V}_{\text{nom}}$, $\mathbf{X}(0)$ in (2.4) is given by

$$\mathbf{X}(0) = \mathbf{X}^{V_{\text{nom}}}(0) = V_{\text{nom}}\mathbf{x}_0 = \mathcal{V}_{\text{nom}}n_A\mathbf{x}_0.$$

Next, we define the limiting propensity functions [29],

$$\bar{a}_j(\mathbf{x}) = \lim_{V \rightarrow \infty} a_j^V(V\mathbf{x})/V, \quad j = 1, \dots, M.$$

For example, if the j th reaction is as in (2.7),

$$a_j^V(\mathbf{x}) = \frac{k_j}{V} x_m x_n \quad \text{and} \quad \bar{a}_j(\mathbf{x}) = k_j x_m x_n.$$

On the other hand, if the j th reaction is of the form (2.8),

$$a_j^V(\mathbf{x}) = \frac{k_j}{2V} x_m (x_m - 1) \quad \text{and} \quad \bar{a}_j(\mathbf{x}) = \frac{1}{2} k_j x_m^2.$$

To describe the thermodynamic limit, we consider the concentration-based state vector $\mathbf{Z}^V(t) = \mathbf{X}^V(t)/V$. In the limit as $V \rightarrow \infty$, $\mathbf{Z}^V(t)$ approaches, almost surely, a

deterministic function $\mathbf{Z}(t)$ that is obtained by solving a system of ODEs known as the system of reaction rate equations (RREs). The theoretical result underpinning this is given in [6, Theorem 2.1 in Chapter 11]. Below, we follow the form of this result as presented in [29]. We also point the reader to [28, Chapter 2] for a detailed exposition of this result.

The concentration vector $\mathbf{Z}^V$ follows the RTC representation [29]

$$(2.10) \quad \mathbf{Z}^V(t) = \mathbf{x}_0 + \sum_{j=1}^M \nu_j V^{-1} Y_j \left(\int_0^t a_j^V(V \mathbf{Z}^V(s)) ds \right).$$

The corresponding system of RREs is described by

$$(2.11) \quad \begin{aligned} \frac{d\mathbf{Z}}{dt} &= F(\mathbf{Z}(t)), \quad t \in [0, T], \\ \mathbf{Z}(0) &= \mathbf{x}_0, \end{aligned}$$

where $F(\mathbf{z}) = \sum_{j=1}^M \nu_j \bar{a}_j(\mathbf{z})$ and $[0, T]$ is the maximal interval of existence of solution for (2.11). The result given in [6, Theorem 2.1 in Chapter 11] (see also [29]), which covers more general classes of Markov processes, states that if for all compact $K \subset \mathbb{R}^N$

$$(2.12) \quad \begin{aligned} \sum_{j=1}^M \|\nu_j\| \sup_{\mathbf{z} \in K} \bar{a}_j(\mathbf{z}) &< \infty, \quad \text{and} \\ F &\text{ is Lipschitz on } K, \end{aligned}$$

then

$$(2.13) \quad \lim_{V \rightarrow \infty} \sup_{s \leq T} \|\mathbf{Z}^V(s) - \mathbf{Z}(s)\| = 0 \quad \text{almost surely.}$$

Note that here $\|\cdot\|$ denotes the Euclidean norm. Therefore, we know that in the limit, as $V \rightarrow \infty$, the stochastic solutions obtained from (2.10) will converge almost surely to the solution of the ODE system (2.11). Note also that both of the conditions in (2.12) hold for the chemical systems under study, because $\bar{a}_j$'s are polynomials.

3. The Next Reaction Method. Several algorithms have been developed for simulating the dynamics of a stochastic chemical reaction network; these include Gillespie's stochastic simulation algorithm (SSA) [8, 11] as well as the Next Reaction Method (NRM) of Gibson and Bruck [7] and its variants [3, 15, 17]. The NRM approach has a number of advantages over the SSA (see [3, section 1] and [20, section III.B], among others): (i) it is cheaper to simulate than the SSA in terms of random numbers generated per iteration; and (ii) it has the ability to handle time-dependent propensity functions and reactions that exhibit delays between initiation and completion. The variant of the NRM that we use below is developed by Anderson in [3], where it is referred to as the *modified next reaction method*.

The NRM simulates RTC dynamics by treating each reaction as an independent stochastic process: indeed, (2.4), (2.5) correspond to a linear combination of Poisson processes with different internal times τ_j , $j = 1, \dots, M$. The approach is then to track the firing of each reaction in terms of these internal times. Given the "current" internal time τ_j , $j = 1, \dots, M$, we denote by τ_j^+ the internal time at which reaction j fires next. At each iteration, the vectors $[\tau_1 \ \tau_2 \ \cdots \ \tau_M]^\top$ and $[\tau_1^+ \ \tau_2^+ \ \cdots \ \tau_M^+]^\top$

Algorithm 3.1 Modified Next Reaction Method [3].

Input: Initial state $\mathbf{X}_0$, final simulation time T , stoichiometric matrix $\boldsymbol{\nu}$, and propensity functions, $\{a_j(\cdot)\}_{j=1}^M$.

Output: A realization of $\mathbf{X}(t, \omega)$.

```

1: % initialization %
2: for  $j = 1, \dots, M$  do
3:   Generate random number  $r_j \sim U(0, 1)$ 
4:    $\tau_j = 0, \tau_j^+ = -\ln(r_j)$ 
5: end for
6:  $t = 0, \mathbf{X}(0) = \mathbf{X}_0$ 
7: % simulation loop %
8: while  $t < T$  do
9:   for  $j = 1, \dots, M$  do
10:    Evaluate  $a_j(\mathbf{X}(t))$  and  $\Delta t_j = \frac{\tau_j^+ - \tau_j}{a_j(\mathbf{X}(t))}$ 
11:   end for
12:   Set  $l = \underset{j}{\operatorname{argmin}} \{\Delta t_j\}_{j=1}^M$ 
13:    $\mathbf{X}(t + \Delta t_l) \leftarrow \mathbf{X}(t) + \boldsymbol{\nu}_l$  {Update state vector}
14:    $t \leftarrow t + \Delta t_l$  {Update global time}
15:   for  $j = 1, \dots, M$  do
16:     $\tau_j \leftarrow \tau_j + a_j \Delta t_l$  {Update internal times of each reaction}
17:   end for
18:   Generate random number  $r_l \sim U(0, 1)$ 
19:    $\tau_l^+ \leftarrow \tau_l^+ - \ln(r_l)$  {Update next reaction time for reaction  $l$ }
20: end while

```

store the current internal time and the next internal time for each reaction. Given these two vectors, one can determine how much physical or global time will elapse before reaction j fires again by considering

$$\Delta t_j = \frac{\tau_j^+ - \tau_j}{a_j(\mathbf{X}(t))}, \quad j = 1, \dots, M.$$

This is a direct consequence of (2.5) and the assumption that a_j remains constant in the interval $[t, t + \Delta t]$ with $\Delta t = \max_j \Delta t_j$. The index of the next reaction to fire is then $l = \operatorname{argmin}(\Delta t_j)$, from which the system state and propensities may be updated and the global time incremented by Δt_l . The next internal time for reaction l to fire is then computed as $\tau_l^+ = \tau_l^+ + \xi$, where ξ represents the duration between events in a Poisson process; the latter implies ξ is exponentially distributed. Each τ_j where $j \neq l$, corresponding to an internal time that has not reached firing, is given the approximate update, $\tau_j = \tau_j + a_j \Delta t_l$, which is discussed in detail in [3, section 4]. An outline of the full NRM algorithm for a general reaction network is given in Algorithm 3.1. We note that $-\ln(r_l)$ is exponentially distributed given that r_l is uniformly distributed in the interval $[0, 1]$.

4. Global sensitivity analysis for stochastic models. In this section, we study convergence of sensitivity indices corresponding to stochastic models to their deterministic counterparts. In section 4.1, we describe the underlying probabilistic setup and global sensitivity analysis via Sobol' indices. In section 4.2, we present a generic result regarding convergence of the Sobol' indices of a family of random

processes. Then, in section 4.3, we show how the generic convergence result can be applied to stochastic chemical systems.

4.1. The basic setup. Stochastic models with uncertain parameters present two sources of uncertainties: intrinsic uncertainty due to stochasticity of the system and uncertainty in model parameters. We denote the probability space carrying intrinsic stochasticity of the system by $(\Omega, \mathcal{F}, \nu)$, where Ω is the sample space equipped with a sigma-algebra $\mathcal{F}$ and a probability measure ν . In stochastic chemical systems, the uncertain model parameters of interest are the reaction rate constants, $k_1, \dots, k_M$. We model these as independent uniformly distributed random variables. Following common practice, we parameterize the uncertainty in the k_i 's using a random vector $\boldsymbol{\theta} = [\theta_1, \dots, \theta_M]^\top$ whose entries are independent $U(-1, 1)$ random variables. For example, if $k_i \sim U(a_i, b_i)$, then $k_i(\theta_i) = \frac{1}{2}(a_i + b_i) + \frac{1}{2}(b_i - a_i)\theta_i$.

The uncertain parameter vector $\boldsymbol{\theta}$ takes values in $\Theta = [-1, 1]^M$. It is convenient to work with the probability space $(\Theta, \mathcal{E}, \lambda)$ for the uncertain parameters, where $\mathcal{E}$ is the Borel sigma-algebra on Θ and λ is the law of $\boldsymbol{\theta}$, $\lambda(d\boldsymbol{\theta}) = 2^{-M}d\boldsymbol{\theta}$. The present setup can be easily extended to cases where the θ_i 's are independent random variables belonging to other suitably chosen distributions. Note also that one can have additional uncertain parameters in a chemical system.

We use Sobol' indices [21, 25, 26] to characterize the sensitivity of a quantity of interest (QoI) to input parameter uncertainties. For example, let $f(\boldsymbol{\theta})$ be a scalar-valued QoI defined in terms of the solution of the RREs corresponding to a chemical system. The first order Sobol' indices corresponding to $f(\boldsymbol{\theta})$ are

$$(4.1) \quad S_j(f) := \frac{\mathbb{V}[\mathbb{E}[f(\boldsymbol{\theta}) \mid \theta_j]]}{\mathbb{V}[f]}, \quad j = 1, \dots, M.$$

These indices quantify the proportion of the QoI variance due to the j th input parameter. Here $\mathbb{E}[f(\boldsymbol{\theta}) \mid \theta_j]$ indicates conditional expectation, and $\mathbb{V}[f]$ denotes the variance of f . For further details on theory and computation methods for Sobol' indices we refer the reader to [21, 24, 25, 26].

4.2. Convergence of stochastic Sobol' indices. We consider a family of stochastic processes $\{f_V(\boldsymbol{\theta}, \omega)\}_{V>0}$ with

$$f_V(\boldsymbol{\theta}, \omega) : \Theta \times \Omega \rightarrow \mathbb{R},$$

which, as discussed below, are assumed to admit a deterministic limit as $V \rightarrow \infty$. The Sobol' indices corresponding to $f_V(\boldsymbol{\theta}, \omega)$ are

$$(4.2) \quad S_j(f_V(\cdot, \omega)) := \frac{\mathbb{V}[\mathbb{E}[f_V(\boldsymbol{\theta}, \omega) \mid \theta_j]]}{\mathbb{V}[f_V(\boldsymbol{\theta}, \omega)]}, \quad j = 1, \dots, M.$$

The following result concerns the convergence of these indices in the limit as $V \rightarrow \infty$.

THEOREM 4.1. *Assume the following.*

1. *There exists $f \in L^2(\Theta, \mathcal{E}, \lambda)$ such that, for almost all $\omega \in \Omega$,*

$$(4.3) \quad f_V(\boldsymbol{\theta}, \omega) \rightarrow f(\boldsymbol{\theta}), \quad \text{as } V \rightarrow \infty, \quad \text{for all } \boldsymbol{\theta} \in \Theta.$$

2. *For almost all $\omega \in \Omega$, $f_V(\boldsymbol{\theta}, \cdot)$ is $\mathcal{E}$ -measurable, and there exists $\varphi_\omega(\boldsymbol{\theta}) \in L^2(\Theta, \mathcal{E}, \lambda)$ such that for all $\boldsymbol{\theta} \in \Theta$,*

$$(4.4) \quad |f_V(\boldsymbol{\theta}, \omega)| \leq \varphi_\omega(\boldsymbol{\theta}) \quad \text{for all } V > 0.$$

Then the stochastic Sobol' indices satisfy

$$S_j(f_V(\cdot, \omega)) \rightarrow S_j(f), \quad \text{as } V \rightarrow \infty, \quad \nu\text{-almost surely.}$$

Proof. By the assumptions of the theorem, there exists a set $F \in \mathcal{F}$ with $\nu(F) = 1$ such that the conditions (4.3) and (4.4) hold for every $\omega \in F$. By (4.4), we observe that $f_V(\theta, \omega) \in L^2(\Theta, \mathcal{E}, \lambda)$ for every $\omega \in F$ and $V > 0$. Thus, we can define the stochastic Sobol' indices (4.2) for $\{f_V(\cdot, \omega)\}_{V>0}$, for every $\omega \in F$.

To show that $f_V(\theta, \omega) \rightarrow f(\theta)$ in $L^2(\Theta, \mathcal{E}, \lambda)$, we note that for every $\omega \in F$ $|f_V(\theta, \omega) - f(\theta)|^2 \rightarrow 0$ pointwise in Θ and

$$|f_V(\theta, \omega) - f(\theta)|^2 \leq 4\varphi_\omega(\theta)^2 \in L^1(\Theta, \mathcal{E}, \lambda).$$

Therefore, invoking the Lebesgue dominated convergence theorem, we have that for all $\omega \in F$, $\int_\Theta |f_V(\theta, \omega) - f(\theta)|^2 \lambda(d\theta) \rightarrow 0$, and thus for every $\omega \in F$

$$\lim_{V \rightarrow \infty} \int_\Theta [f_V(\theta, \omega)]^r \lambda(d\theta) = \int_\Theta [f(\theta)]^r \lambda(d\theta), \quad r = 1, 2.$$

The convergence of the first and second moments of $f_V(\cdot, \omega)$ clearly implies

$$\lim_{V \rightarrow \infty} \mathbb{V}(f_V(\cdot, \omega)) = \mathbb{V}(f(\cdot)) \quad \text{for all } \omega \in F.$$

To finish the proof of the theorem, we need to show

$$\lim_{V \rightarrow \infty} \mathbb{V}\{\mathbb{E}(f_V(\cdot, \omega)|\theta_j)\} = \mathbb{V}\{\mathbb{E}(f(\cdot)|\theta_j)\} \quad \text{for all } \omega \in F, j = 1, \dots, M.$$

Using the reverse triangle inequality and Jensen's inequality, we observe

$$\begin{aligned} \left| \|\mathbb{E}(f_V(\cdot, \omega)|\theta_j)\|_{L^2(\Theta)} - \|\mathbb{E}(f(\cdot)|\theta_j)\|_{L^2(\Theta)} \right| &\leq \|\mathbb{E}(f_V(\cdot, \omega)|\theta_j) - \mathbb{E}(f(\cdot)|\theta_j)\|_{L^2(\Theta)} \\ &= \|\mathbb{E}(f_V(\cdot, \omega) - f(\cdot)|\theta_j)\|_{L^2(\Theta)} \\ &\leq \|f_V(\cdot, \omega) - f(\cdot)\|_{L^2(\Theta)}, \end{aligned}$$

and thus, for all $\omega \in F$

$$\lim_{V \rightarrow \infty} \|\mathbb{E}(f_V(\cdot, \omega)|\theta_j)\|_{L^2(\Theta)} = \|\mathbb{E}(f(\cdot)|\theta_j)\|_{L^2(\Theta)}.$$

Since

$$\begin{aligned} \mathbb{V}\{\mathbb{E}(f_V(\cdot, \omega)|\theta_j)\} &= \mathbb{E}\{\mathbb{E}(f_V(\cdot, \omega)|\theta_j)^2\} - \mathbb{E}\{\mathbb{E}(f_V(\cdot, \omega)|\theta_j)\}^2 \\ &= \|\mathbb{E}(f_V(\cdot, \omega)|\theta_j)\|_{L^2(\Theta)}^2 - \mathbb{E}\{f_V(\cdot, \omega)\}^2, \end{aligned}$$

we have, for all $\omega \in F$,

$$(4.5) \quad \lim_{V \rightarrow \infty} \mathbb{V}\{\mathbb{E}(f_V(\cdot, \omega)|\theta_j)\} = \|\mathbb{E}(f(\cdot)|\theta_j)\|_{L^2(\Theta)}^2 - \mathbb{E}\{f(\cdot)\}^2 = \mathbb{V}\{\mathbb{E}(f(\cdot)|\theta_j)\}.$$

This, along with the convergence of the (unconditional) variance implies

$$\lim_{V \rightarrow \infty} S_j(f_V(\cdot, \omega)) = \lim_{V \rightarrow \infty} \frac{\mathbb{V}\{\mathbb{E}(f_V(\theta, \omega)|\theta_j)\}}{\mathbb{V}\{f_V(\theta, \omega)\}} = \frac{\mathbb{V}\{\mathbb{E}(f(\theta)|\theta_j)\}}{\mathbb{V}\{f\}} = S_j(f)$$

for all $\omega \in F, j = 1, \dots, M$. $\square$

Remark 4.2. A slight modification of the proof of Theorem 4.1 leads to a more general result: namely, we can obtain almost sure convergence of the indices,

$$(4.6) \quad S_U(f_V(\cdot, \omega)) := \frac{\mathbb{V}[\mathbb{E}[f_V(\cdot, \omega) \mid \boldsymbol{\theta}_U]]}{\mathbb{V}[f_V(\cdot, \omega)]},$$

where $U = \{j_1, j_2, \dots, j_s\} \subseteq \{1, 2, \dots, M\}$ and $\boldsymbol{\theta}_U = [\theta_{j_1} \ \theta_{j_2} \ \dots \ \theta_{j_s}]^\top$, to $S_U(f(\cdot))$.

We recall the *total* Sobol' indices [21]

$$(4.7) \quad T_j(f_V(\cdot, \omega)) := \sum_{U \ni j} S_U(f_V(\cdot, \omega)), \quad j = 1, \dots, M.$$

These indices quantify the relative contribution of θ_j by itself, and through its interactions with the other coordinates of $\boldsymbol{\theta}$, to the variance of $f_V(\cdot, \omega)$. In view of Remark 4.2, under the conditions of Theorem 4.1

$$\lim_{V \rightarrow \infty} T_j(f_V(\cdot, \omega)) = T_j(f(\cdot)) \quad \text{for almost all } \omega \in \Omega, \quad j = 1, \dots, M.$$

4.3. Application to stochastic chemical kinetics. Consider the (concentration based) state vector $\mathbf{Z}^V(t, \boldsymbol{\theta}, \omega)$ of a stochastic chemical system and its deterministic counterpart $\mathbf{Z}(t, \boldsymbol{\theta})$, corresponding the thermodynamic limit. Recall that $\boldsymbol{\theta} \in \Theta$ parameterizes the uncertainty in reaction rate constants. In the present work, we focus on a scalar time-independent QoI $G(\mathbf{Z}^V(t, \boldsymbol{\theta}, \omega))$ and its deterministic counterpart $G(\mathbf{Z}(t, \boldsymbol{\theta}))$. Specifically, G takes a vector function $\mathbf{z}(t)$ and returns a scalar QoI. Examples include

$$(4.8a) \quad G(\mathbf{z}(t)) = z_i(t^*) \quad \text{for fixed } t^* \in [0, T] \text{ and } i \in \{1, \dots, N\}, \quad \text{or}$$

$$(4.8b) \quad G(\mathbf{z}(t)) = \frac{1}{T} \int_0^T z_i(t) dt \quad \text{for a fixed } i \in \{1, \dots, N\}.$$

In general, we assume $G : L^\infty([0, T]; \mathbb{R}^N) \rightarrow \mathbb{R}$ to be a continuous function. Note that $L^\infty([0, T]; \mathbb{R}^N)$ is equipped with norm $\|\cdot\|_\infty$ given by $\|\mathbf{z}\|_\infty = \sup_{t \in [0, T]} \|\mathbf{z}(t)\|$, where as before $\|\cdot\|$ denotes the Euclidean vector norm.

To put things in the notation of the previous subsection, we consider

$$f_V(\boldsymbol{\theta}, \omega) = G(\mathbf{Z}^V(t, \boldsymbol{\theta}, \omega)), \quad \boldsymbol{\theta} \in \Theta, \omega \in \Omega,$$

and the corresponding limiting (deterministic) quantity, $f(\boldsymbol{\theta}) = G(\mathbf{Z}(t, \boldsymbol{\theta}))$. Note that by (2.13), for fixed $\boldsymbol{\theta} \in \Theta$, as $V \rightarrow \infty$

$$\|\mathbf{Z}^V(\cdot, \boldsymbol{\theta}, \omega) - \mathbf{Z}(\cdot, \boldsymbol{\theta})\|_\infty \rightarrow 0 \quad \text{for almost all } \omega \in \Omega.$$

Therefore, by the continuous mapping theorem (see, e.g., [5]), for each $\boldsymbol{\theta} \in \Theta$,

$$(4.9) \quad f_V(\boldsymbol{\theta}, \omega) \rightarrow f(\boldsymbol{\theta}), \quad \text{almost surely,}$$

as $V \rightarrow \infty$. We consider the convergence of the stochastic Sobol' indices $S_j(f_V(\cdot, \omega))$ to their deterministic counterparts $S_j(f(\cdot))$, $j = 1, \dots, M$, as $V \rightarrow \infty$, i.e., in the thermodynamic limit. Here we discuss how things can be put in the framework of Theorem 4.1, which would then imply almost sure convergence of the stochastic Sobol' indices to their limiting deterministic counterparts.

Theorem 4.1 requires existence of a set of full measure in Ω such that the convergence in (4.9) holds. To ensure this, we consider a modification of $f_V(\boldsymbol{\theta}, \omega)$ as follows. We know that for each $\boldsymbol{\theta} \in \Theta$, there exists a set of full measure $F_\theta \subseteq \Omega$ for which the convergence (4.9) holds. Define

$$\tilde{f}_V(\boldsymbol{\theta}, \omega) = \begin{cases} f_V(\boldsymbol{\theta}, \omega) & \text{if } \omega \in F_\theta, \\ f(\boldsymbol{\theta}) & \text{otherwise.} \end{cases}$$

Note that we have $\nu(\{\omega \in \Omega : \tilde{f}_V(\boldsymbol{\theta}, \cdot) = f_V(\boldsymbol{\theta}, \omega)\}) = 1$ for every $\boldsymbol{\theta} \in \Theta$. That is, $\tilde{f}_V(\boldsymbol{\theta}, \cdot)$ is a modification of $f_V(\boldsymbol{\theta}, \cdot)$. Note that this modification satisfies the following: for every $\omega \in \Omega$, $\tilde{f}_V(\boldsymbol{\theta}, \omega) \rightarrow f(\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \Theta$. With a slight abuse of notation, we will denote this modification by $f_V(\boldsymbol{\theta}, \omega)$ from this point on. To ensure that Theorem 4.1 applies, we also need the boundedness assumption (4.4).

To discuss the boundedness assumption (4.4), we take a step back and first discuss conditions ensuring boundedness of the stochastic system trajectory $\{\mathbf{Z}^V(t, \boldsymbol{\theta}, \omega)\}_{V>0}$. Consider the state vector $\mathbf{X}^V(t)$. Nonnegativity of this state vector requires the propensity functions to be proper [19]: for $j = 1, \dots, M$, we assume, for all $\mathbf{x} \in \mathbb{Z}_+^N$, if $\mathbf{x} + \boldsymbol{\nu}_j \notin \mathbb{Z}_+^N$, then $a_j^V(\mathbf{x}) = 0$. Boundedness of components of $\mathbf{X}^V(t)$ requires further (mild) assumptions, as formalized in [19, Theorem 2.8 and 2.11]. Interestingly, the only requirements concern the stoichiometric matrix $\boldsymbol{\nu}$. Namely, assuming the existence of a vector $\boldsymbol{\alpha} \in \mathbb{Z}_{\geq 0}^N$ such that $\boldsymbol{\alpha}^\top \boldsymbol{\nu} \leq 0$ and $\alpha_i > 0$ is necessary and sufficient for boundedness of $X_i^V(t)$. Specifically, if such an $\boldsymbol{\alpha}$ exists, $\boldsymbol{\alpha}^\top \mathbf{X}^V(t) = \boldsymbol{\alpha}^\top (\mathbf{X}^V(0) + \boldsymbol{\nu} \mathbf{R}(t)) \leq \boldsymbol{\alpha}^\top \mathbf{X}^V(0)$. Therefore,

$$X_i^V(t) \leq (1/\alpha_i) \boldsymbol{\alpha}^\top \mathbf{X}^V(0) = (V/\alpha_i) \boldsymbol{\alpha}^\top \mathbf{x}_0.$$

Thus, in terms of concentrations

$$Z_i^V(t) = X_i^V/V \leq (1/\alpha_i) \boldsymbol{\alpha}^\top \mathbf{x}_0.$$

Therefore, we have that the i th component of $\mathbf{Z}^V$ remains uniformly bounded by $(1/\alpha_i) \boldsymbol{\alpha}^\top \mathbf{x}_0$. Moreover, this bound is independent of the reaction rate constants, i.e., independent of $\boldsymbol{\theta}$. Thus, if a vector $\boldsymbol{\alpha}$ satisfying the aforementioned properties exists for all the components of the state vector, then the concentration based state vector $\mathbf{Z}^V$ remains uniformly bounded by a constant. In fact, we need to only ensure boundedness of the components of $\mathbf{Z}^V$ that appear in definition of G . Given the function G , which defines the QoI, is sufficiently well behaved, one may argue that f_V inherits the boundedness necessary to satisfy (4.4). For example, if G is defined as in (4.8), then establishing boundedness of $\{Z_i^V(t, \boldsymbol{\theta}, \omega)\}_{V>0}$ is sufficient to satisfy (4.4) for the QoI, f_V .

5. Numerical results. In light of the convergence properties exhibited by stochastic chemical reaction systems, we aim to demonstrate numerically the results of Theorem 4.1. Convergence results will be presented first for the Michaelis–Menten reaction system, followed by an application of Theorem 4.1 to the task of dimension reduction, considering a higher-dimensional example arising from the study of genetic networks. Attention will also be devoted to the computation of Sobol' indices and the random sampling necessary to compute the stochastic Sobol' indices introduced in section 4.

5.1. The Michaelis–Menten system. The Michaelis–Menten reaction is the most well-known example of enzymatic catalysis in the chemical kinetics literature [2, 11, 15]:



In (5.1), the substrate S binds to the enzyme E to form the complex C . The complex may either dissociate back into the substrate and enzyme or dissociate into the enzyme and a product P . Figure 5.1 depicts 25 realizations of the reaction dynamics using the NRM algorithm with a final time of $T = 50$. The parameters, corresponding to the rate constants in the propensity functions, are fixed to the nominal values $\bar{k}_1 = 10^6$, $\bar{k}_2 = 10^{-4}$, and $\bar{k}_3 = 0.1$ provided in [30]. Figure 5.1 depicts concentrations of each species for a system size of $V_{\text{nom}} = n_A \mathcal{V}_{\text{nom}}$, where the nominal volume of the reaction system is $\mathcal{V}_{\text{nom}} = 10^{-15} \text{ m}^3$.

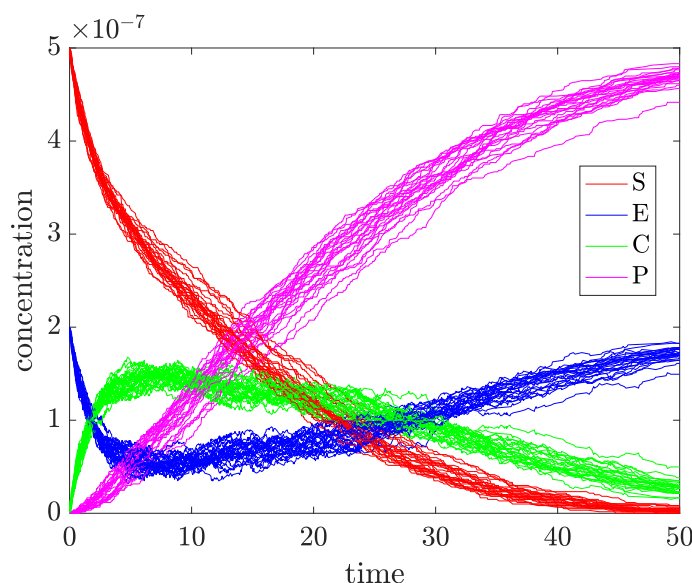


FIG. 5.1. 25 realizations of Michaelis–Menten trajectories computed via NRM with nominal parameters, varying ω .

In Figure 5.2 we illustrate convergence of the RTC trajectories to the RRE trajectories as the system size increases. We hold the parameters fixed to their nominal values and plot 25 realizations of the product $P^V(t, \omega) = Z_4^V(t, \omega)$ along with the corresponding RRE trajectory. As the system size increases, the ensemble of RTC trajectories converges to the RRE trajectory. In Figure 5.2, the quantity m denotes the multiplicative factor by which the system size is varied. For the purpose of the simulation, m is related to the system size by the relation $V = mV_{\text{nom}}$.

5.1.1. The QoI. In the present study we focus on the stochastic QoI

$$f_V(\theta, \omega) = \frac{1}{T} \int_0^T Z_4^V(t; \theta, \omega) dt,$$

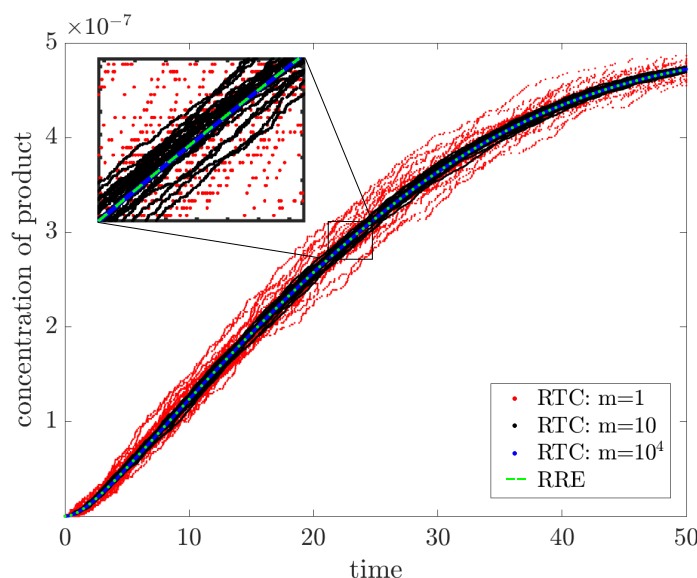


FIG. 5.2. Convergence of the product $P^V(t, \omega)$ to the corresponding RRE solution at the nominal parameter values plotted as system size grows.

where $\mathbf{Z}^V$ is the solution of the RTC representation. The corresponding deterministic QoI is

$$f(\boldsymbol{\theta}) = \frac{1}{T} \int_0^T Z_4(t; \boldsymbol{\theta}) dt,$$

where $\mathbf{Z}$ is computed by solving the accompanying RRE. To get a sense of the statistical properties of the QoI, we sample f_V and f over the uncertain parameter domain given by $\Theta = [-1, 1]^3$, and with the uncertain rate constants defined as

$$k_i(\theta_i) = \bar{k}_i + (0.1\bar{k}_i)\theta_i, \quad i = 1, 2, 3,$$

where $\bar{k}_i$'s are the nominal reaction rate constants as defined above. Figure 5.3 shows probability density functions (PDFs) of f sampled in Θ , f_V sampled in $\Theta \times \Omega$, and f_V sampled in Ω while using nominal parameters. All samples of f_V used in Figure 5.3 use the $V = V_{\text{nom}}$.

5.1.2. Global sensitivity analysis. In this section, we turn to estimating Sobol' indices in both the stochastic and deterministic settings. For the purpose of this study, we focus on the computation and convergence of the total Sobol' indices. The method detailed below can be applied to Sobol' indices of any order.

Sobol' indices measure the relative contribution of a subset of uncertain parameters to the variance of some QoI. Consequently, it is natural to consider QoIs that are deterministic functions of these uncertain parameters, without any additional variance contributed by a secondary source. When modeling chemical systems using stochastic processes, such as the RTC representation, the model parameters and internal stochasticity both provide sources of uncertainty, which must be accounted for separately. We summarize the process of estimating Sobol' indices in the deterministic and stochastic cases in the Algorithm 5.1, where the number of uncertain parameters is denoted p . Note that it is not always the case that $p = M$, the number of reactions.

In the stochastic setting, fixing a particular ω_i turns f_V into a deterministic function of the uncertain parameters. From that point, the process of estimating

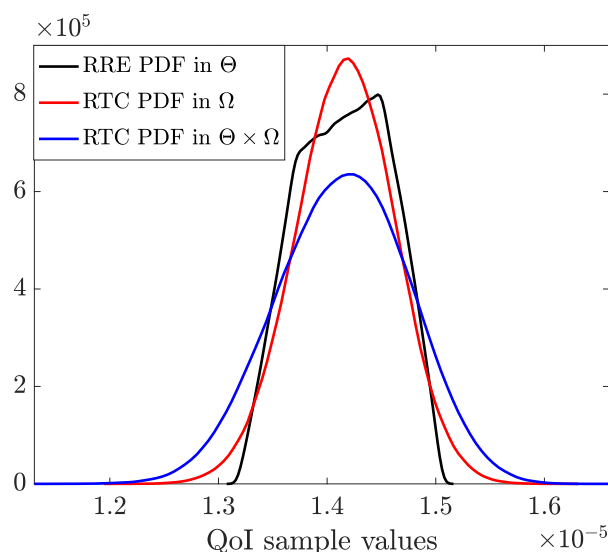


FIG. 5.3. Estimated PDFs of f_V sampled over Ω and $\Theta \times \Omega$ and f sampled over Θ , respectively.

Algorithm 5.1 Sobol' indices for a chemical system with fixed system size.

Input: Method of evaluating $f_V(\boldsymbol{\theta}, \omega)$ and $f(\boldsymbol{\theta})$, N_s : number of parameter samples, set of M_s random seeds $\{\xi_i\}_{i=1}^{M_s}$, system size V .

Output: Total Sobol' indices: $\{T_1^V(\omega_i), \dots, T_p^V(\omega_i)\}_{i=1}^{M_s}$ and $\{T_1, \dots, T_p\}$.

```

1: Draw  $N_s(p+2)$  samples uniformly in  $\Theta$  {see [24] for details}
2: % stochastic indices %
3: for  $i = 1, \dots, M_s$  do
4:   Seed random number generator with  $\xi_i$ , corresponding to realization  $\omega_i$ 
5:   for  $j = 1, \dots, N_s(p+2)$  do
6:     Evaluate and store  $f_V(\boldsymbol{\theta}_j, \omega_i)$  samples
7:   end for
8:   Using  $f_V$  samples, estimate Sobol' indices:  $\{T_1^V(\omega_i), \dots, T_p^V(\omega_i)\}$ 
9: end for
10: % deterministic indices %
11: for  $j = 1, \dots, N_s(p+2)$  do
12:   Evaluate and store  $f(\boldsymbol{\theta}_j)$  samples
13: end for
14: Using  $f$  samples, estimate Sobol' indices:  $\{T_1, \dots, T_p\}$ 

```

Sobol' indices is identical to the deterministic case. We estimate Sobol' indices using Monte Carlo integration; see [21, 24] or [22, section 4.5] for details. In Algorithm 5.1, the cost of estimating first order and total indices for each fixed ω_i is $N_s(p+2)$ evaluations of the QoI, where N_s is user-defined.

The realizations of the stochastic indices correspond to $\omega_i \in \Omega$, $i = 1, \dots, M_s$, prescribed by the choice of random seed. We also note that the stochastic indices are functions of the given system size, while the deterministic indices do not depend on V and should not be recomputed each time V is changed. For a fixed V , we may compare the distribution of each T_i^V with the deterministic value of T_i .

Returning to the Michaelis–Menten example, in Figure 5.4 we plot the PDFs of the stochastic total indices corresponding to the default V , where $m = 1$.

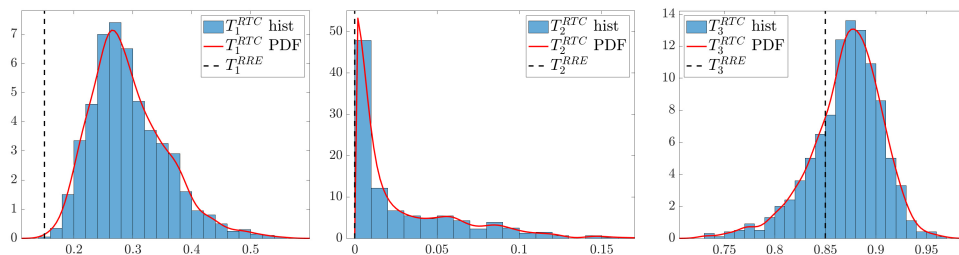


FIG. 5.4. Histogram and PDF estimates for the total Sobol' indices for k_1 , k_2 , and k_3 , respectively. Black dashed lines indicate the deterministic value of the RRE total indices.

The deterministic indices, estimated with $N_s = 10^7$ samples, are $T_1 \approx 1.5 \times 10^{-1}$, $T_2 \approx 1.2 \times 10^{-7}$, and $T_3 \approx 8.5 \times 10^{-1}$, indicating that the third reaction, where the complex dissociates into the enzyme and the product, is the most important, and the second reaction, where the complex dissociates into the enzyme and substrate, is the least important, contributing almost no variance.

5.1.3. Convergence of Sobol' indices. One may verify that the conditions on the QoI necessary for Theorem 4.1 to hold are satisfied in the present case. Thus we demonstrate numerically the convergence of the stochastic Sobol' indices to the stated deterministic values. After we have computed multiple realizations of the stochastic indices at increasing, discrete values of V , we examine the evolution of their distribution as V increases.

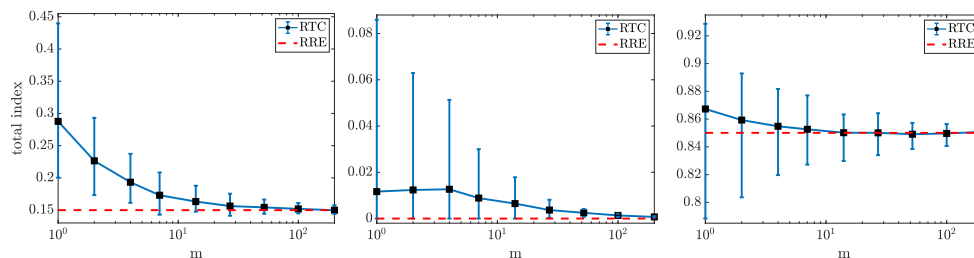


FIG. 5.5. Convergence of the mean total Sobol' index as a function of V for parameters k_1 , k_2 , and k_3 , respectively. Note that the vertical axes of each figure are not over the same range. The lower and upper bounds of the error bars indicate the 5th and 95th percentiles, respectively.

Figure 5.5 demonstrates the convergence of $\mathbb{E}[T_i^{V_m}(\omega)]$ for $i = 1, 2, 3$, for increasing values of system size $V_m = mV_{\text{nom}}$, $m = 1, \dots, 200$. The error bars represent the 5th and 95th percentiles of the distribution of stochastic indices at a particular system size, where $M_s = 100$ different values of ω are sampled to construct the distribution for each discrete value of V . Figure 5.5 suggests the convergence of the PDF for each $T_i^V(\omega)$ to a Dirac distribution centered at the deterministic value of the Sobol' index corresponding to the RRE. This sort of convergence may also be demonstrated for lower order Sobol' indices, as addressed in Remark 4.2.

Figure 5.6 gives a three-dimensional view of the convergence in Figure 5.5. We plot a series of normalized histograms at specific values of m , converging to Dirac distributions centered at the RRE total indices. These histograms, even for two

orders of magnitude difference in V , show a clear trend towards the limiting values given by the RRE.

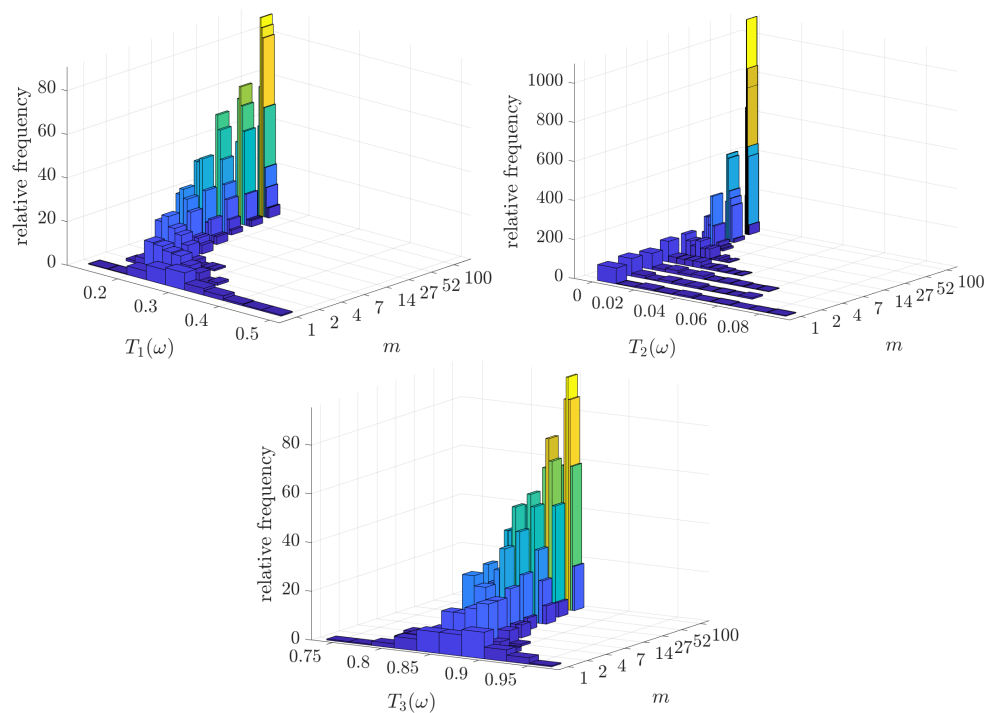


FIG. 5.6. Histograms at discrete V values of the total Sobol' indices for k_1 , k_2 , and k_3 , respectively. The vertical axes represent the relative frequency of the indices due to normalized histograms.

Figures 5.5 and 5.6 can perhaps most naturally be understood as illustrating the convergence in distribution of the RTC Sobol' indices, an implication of the pointwise convergence of the PDF. In this case, $T_i^V(\omega)$ is the random variable that converges in distribution for each $i = 1, 2, 3$ as V approaches infinity.

While the analysis of convergence rates is beyond the scope of the present paper, Figure 5.7 provides a preliminary result in that direction by displaying the variance of the stochastic total Sobol' indices for increasing values of m . As Figures 5.5 and 5.6 indicate, the variance of the total indices approaches zero as the system size approaches infinity. Figure 5.7 indicates that this convergence occurs with a rate of $O(1/V)$. Here the sample variance is estimated with 100 realizations of the stochastic total Sobol' indices. We hypothesize that the faster decay of the variance of $T_2(\omega)$ is due to its small size in the thermodynamic limit.

5.2. The genetic oscillator system. Returning to the original question illustrated in Figure 1.1, we aim to use the sensitivity information from a deterministic chemical model to infer the sensitivities of its stochastic counterpart. The goal is to perform well-informed dimension reduction on the expensive stochastic model, while only requiring samples from the cheaper deterministic model. To perform meaningful dimension reduction, here we consider a higher dimensional model than previously considered. We consider the genetic oscillator system presented in [27], which models the evolution of activator and repressor proteins that govern the circadian clocks of a wide variety of organisms. The system consists of nine species, including genes,

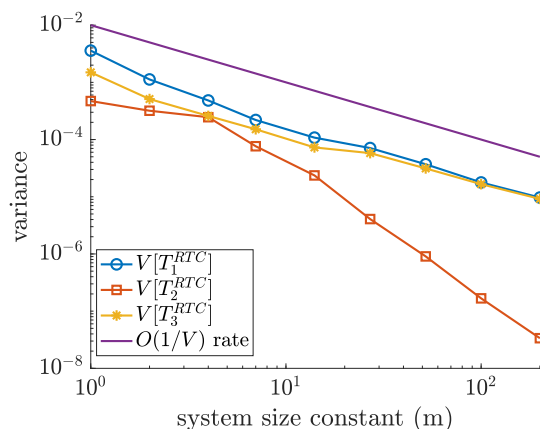


FIG. 5.7. Log-log plot of the rate of convergence of the second moment for each total index.

TABLE 5.1

Genetic oscillator reactions, propensity functions, and nominal parameter values; see [23].

Reaction	Propensity function	Parameter	Value
$P_a \rightarrow P_a + mRNA_a$	$\alpha_A P_a$	α_A	50.0
$P_a - A \rightarrow P_a - A + mRNA_a$	$\alpha_a \alpha_A P_a - A$	α_R	0.01
$P_r \rightarrow P_r + mRNA_r$	$\alpha_R P_r$	β_A	50.0
$P_r - A \rightarrow P_r - A + mRNA_r$	$\alpha_r \alpha_R P_r - A$	β_R	5.0
$mRNA_a \rightarrow mRNA_a + A$	$\beta_A mRNA_a$	γ_C	20.0
$mRNA_r \rightarrow mRNA_r + R$	$\beta_R mRNA_r$	γ_A	1.0
$A + R \rightarrow C$	$\gamma_C AR$	θ_A	50.0
$P_a + A \rightarrow P_a - A$	$\gamma_A P_a A$	γ_R	1.0
$P_a - A \rightarrow P_a + A$	$\theta_A P_a - A$	θ_R	1.0
$P_r + A \rightarrow P_r - A$	$\gamma_R P_r A$	δ_A	1.0
$P_r - A \rightarrow P_r + A$	$\theta_R P_r - A$	δ_R	0.2
$A \rightarrow \emptyset$	$\delta_A A$	δ_{MA}	10.0
$R \rightarrow \emptyset$	$\delta_R R$	δ_{MR}	0.5
$mRNA_a \rightarrow \emptyset$	$\delta_{MA} mRNA_a$	δ'_A	1.0
$mRNA_r \rightarrow \emptyset$	$\delta_{MR} mRNA_r$	α_a	10.0
$C \rightarrow R$	$\delta'_A C$	α_r	5000

mRNAs, and the two proteins. We have $M = 16$ reactions and sixteen uncertain parameters. Following the form of the chemical system presented in [23], we provide the reaction diagrams, propensity functions, and nominal parameter values in Table 5.1.

As with the Michaelis–Menten system, the RTC representation models the evolution of the stochastic system, and the RRE models the deterministic system, with the two models linked by the thermodynamic limiting process. Figure 5.8 shows a sample trajectory of the stochastic system, simulated via the NRM. In 5.8, all parameters are set to nominal values and the only nonzero initial states are P_a and P_r , with one molecule of each. We plot the activator protein A , the repressor protein R , and the complex C up to final time $T = 50$. We then will use the sensitivity information gained from the cheaper, deterministic model (RRE) to make conclusions about parameter importance in the more expensive, stochastic model (RTC).

We define the stochastic and deterministic QoIs, respectively, as

$$f_V(\boldsymbol{\theta}, \omega) = \frac{1}{T} \int_0^T R^V(t; \boldsymbol{\theta}, \omega) dt \quad \text{and} \quad f(\boldsymbol{\theta}) = \frac{1}{T} \int_0^T R(t; \boldsymbol{\theta}) dt,$$

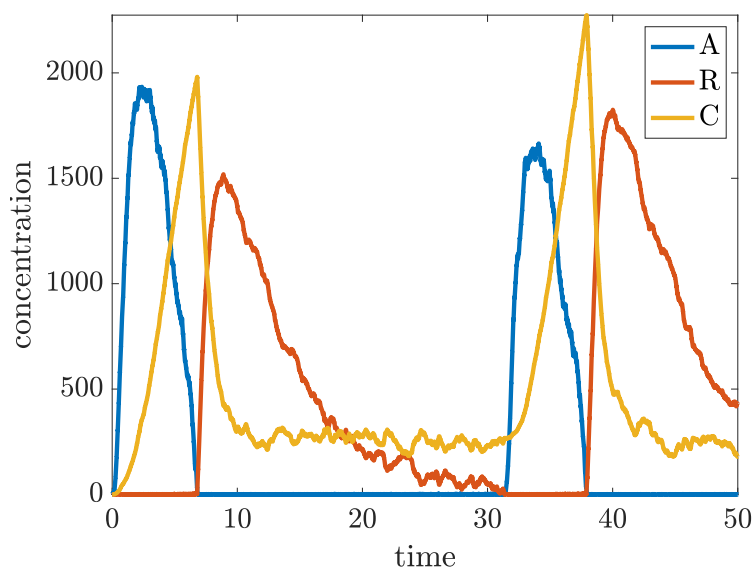


FIG. 5.8. Trajectories of the three dominant species at nominal parameters via the NRM.

where R^V is the concentration of the repressor computed via the NRM and R is the concentration of the repressor computed as the solution to the accompanying RRE. Using the Monte Carlo method presented in [21, 22], we then estimate the total Sobol' indices for the deterministic model. Figure 5.9 shows the total Sobol' indices. It is clear that $\alpha_A, \beta_A, \delta_{MA}$, and α_a are the four most important parameters, capturing over 50% of the variance of the deterministic QoI.

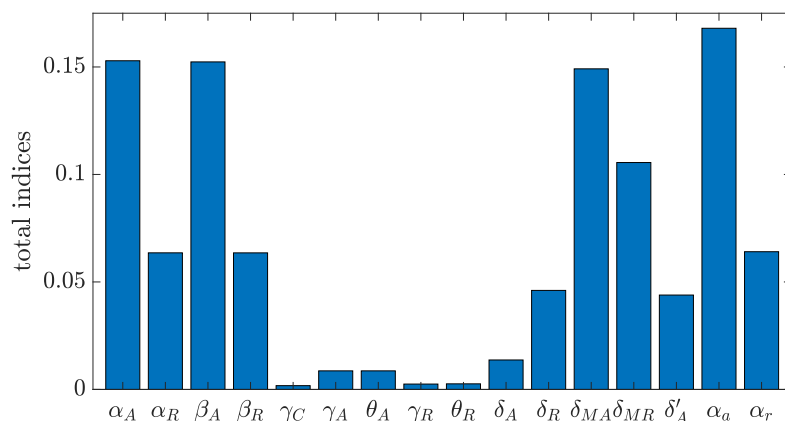


FIG. 5.9. Estimated total Sobol' indices for the genetic oscillator RRE.

We can determine unimportant inputs by putting an importance threshold on the total Sobol' indices; parameters whose Sobol' index falls below the threshold will be considered unimportant. For instance, using 0.02 as a threshold, we identify $\gamma_C, \gamma_A, \theta_A, \gamma_R, \theta_R$, and δ_A as the six least important parameters, capturing less than 5% of the variance of the deterministic QoI. We then propose a reduced-dimensional model, where the six least important parameters are fixed at their nominal values,

reducing the dimensionality from sixteen to ten. To verify that this lower-dimensional model remains an accurate representation of the full model, we sample the stochastic QoI and plot its PDF while fixing and varying the unimportant parameters; see Figure 5.10. The red dashed line, corresponding to the reduced model with the six least important parameters fixed, has a negligible difference with the PDF of the full model. Increasing the threshold from 0.02 to 0.05 adds δ_R and δ'_A to the unimportant category. However, as seen in Figure 5.10, the PDF of the resulting reduced model (dashed green line), obtained by fixing now eight parameters, shows a notable difference from the PDF of the full model. This illustrates the balance one must strike between fixing unimportant parameters to reduce parameter dimension and the loss of information that may result from using a cheaper model. Finally, we illustrate the impact of fixing the four most important parameters (black dashed line in Figure 5.10). This approach fixes every parameter with a total Sobol' index greater than 0.15 ($\alpha_A, \beta_A, \delta_{MA}$, and α_a). This results in a substantial underestimation of the variance and a potential loss of valuable model information.

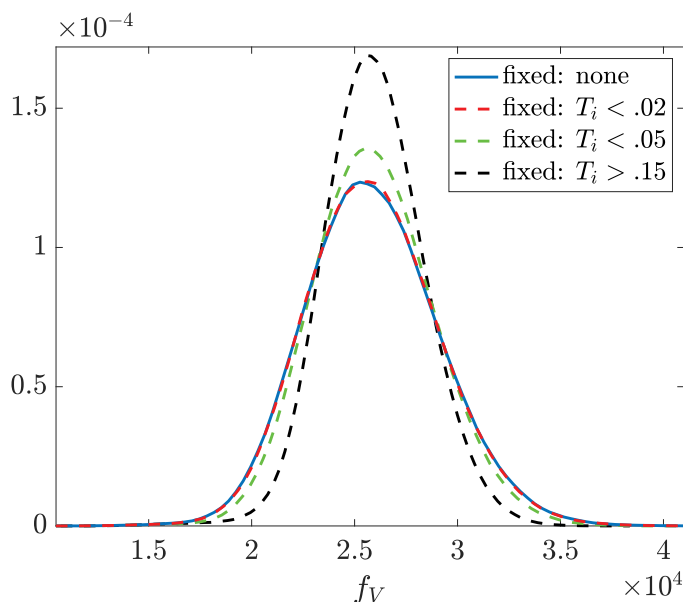


FIG. 5.10. PDFs of the stochastic QoI, f_V , sampled while fixing the following parameters: Black line ($\alpha_A, \beta_A, \delta_{MA}, \alpha_a$), green line ($\gamma_C, \gamma_A, \theta_A, \gamma_R, \theta_R, \delta_A, \delta_R, \delta'_A$), red line ($\gamma_C, \gamma_A, \theta_A, \gamma_R, \theta_R, \delta_A$), black line without fixed parameters. Total index thresholds are provided for each PDF.

6. Conclusions. Sensitivity analysis is often performed on simplified surrogate models with the *hope* that (1.1) holds, i.e., the hope, explicit or not, that the results from the analysis of a surrogate model will hold for the full model. We have presented here a partial result in that direction showing this assertion to be true for a specific class of problems (chemical systems), a specific type of surrogate (obtained from the thermodynamic limit), and a specific GSA approach (Sobol' indices). Our study not only shows and justifies, in an arguably restricted framework, that GSA can sometimes be done “on the cheap;” it also, we argue, reflects important properties of the GSA methods themselves.

An immediate direction for future work is to complement the presented theo-

retical framework by deriving results on the rate of convergence of the stochastic indices to those of the RREs. Moreover, further study should consider other types of limiting processes linking surrogates and full models such as homogenization of differential equations, discretization, and projections, as well as more general types of GSA methods.

Acknowledgments. We acknowledge the computing resources provided on Henry2, a high-performance computing cluster operated by North Carolina State University. We also thank Andrew Petersen for his assistance with distributed and parallel computations, which was provided through the Office of Information Technology HPC services at NC State.

REFERENCES

- [1] D. F. ANDERSON, D. J. HIGHAM, S. C. LEITE, AND R. J. WILLIAMS, *On constrained Langevin equations and (bio)chemical reaction networks*, Multiscale Model. Simul., 17 (2019), pp. 1–30, <https://doi.org/10.1137/18M1190999>.
- [2] D. ANDERSON AND T. KURTZ, *Continuous time Markov chain models for chemical reaction networks*, in Design and Analysis of Biomolecular Circuits: Engineering Approaches to Systems and Synthetic Biology, H. Koepl, G. Setti, M. di Bernardo, and D. Densmore, eds., Springer, New York, 2011, pp. 3–42.
- [3] D. F. ANDERSON, *A modified next reaction method for simulating chemical systems with time dependent propensities and delays*, J. Chem. Phys., 127 (2007), 214107.
- [4] D. F. ANDERSON AND D. J. HIGHAM, *Multilevel Monte Carlo for continuous time Markov chains, with applications in biochemical kinetics*, Multiscale Model. Simul., 10 (2012), pp. 146–179, <https://doi.org/10.1137/110840546>.
- [5] R. DURRETT, *Probability Theory and Examples*, Cambridge University Press, Cambridge, UK, 2019.
- [6] S. ETHIER AND T. KURTZ, *Markov Processes: Characterization and Convergence*, John Wiley & Sons, New York, 1986.
- [7] M. GIBSON AND J. BRUCK, *Efficient exact stochastic simulation of chemical systems with many species and many channels*, J. Phys. Chem. A, 104 (2000), pp. 1876–1889.
- [8] D. GILLESPIE, *Stochastic simulation of chemical kinetics*, Annu. Rev. Phys. Chem., 58 (2007), pp. 35–55.
- [9] D. T. GILLESPIE, *A general method for numerically simulating the stochastic time evolution of coupled chemical reactions*, J. Comput. Phys., 22 (1976), pp. 403–434.
- [10] J. L. HART, A. ALEXANDERIAN, AND P. A. GREMAUD, *Efficient computation of Sobol’ indices for stochastic models*, SIAM J. Sci. Comput., 39 (2017), pp. A1514–A1530, <https://doi.org/10.1137/16M106193X>.
- [11] D. J. HIGHAM, *Modeling and simulating chemical reactions*, SIAM Rev., 50 (2008), pp. 347–368, <https://doi.org/10.1137/060666457>.
- [12] B. IOOSS AND P. LEMAITRE, *A review on global analysis methods*, in Uncertainty Management in Simulation-Optimization of Complex Systems, G. Dellino and C. Meloni, eds., Springer, New York, 2015, pp. 101–122.
- [13] A. JANON, M. NODÉ, AND C. PRIEUR, *Uncertainties assessment in global sensitivity indices estimation from metamodels*, Int. J. Uncertain. Quantif., 4 (2104), pp. 21–36.
- [14] S. KOZIEL, D. ECHEVERRIA CIAURRI, AND L. LEIFSSON, *Surrogate-based methods*, in Computational Optimization, Methods and Algorithms, S. Koziel and X. Yang, eds., Stud. Comput. Intell. 356, Springer, Berlin, 2011, pp. 33–59.
- [15] O. LE MAÎTRE, O. KNIO, AND A. MORAES, *Variance decomposition in stochastic simulators*, J. Chem. Phys., 142 (2015), 06B620.1.
- [16] O. P. LE MAÎTRE AND O. M. KNIO, *PC analysis of stochastic differential equations driven by Wiener noise*, Reliab. Eng. Syst. Safe., 135 (2015), pp. 107–124.
- [17] M. NAVARRO JIMENEZ, O. LE MAÎTRE, AND O. KNIO, *Global sensitivity analysis in stochastic simulators of uncertain reaction networks*, J. Chem. Phys., 145 (2016), 244106.
- [18] E. QIAN, B. PEHERSTORFER, D. O’MALLEY, V. VESSELINOV, AND K. WILLCOX, *Multifidelity Monte Carlo estimation of variance and sensitivity indices*, SIAM/ASA J. Uncertain. Quantif., 6 (2018), pp. 683–706, <https://doi.org/10.1137/17M1151006>.

- [19] M. RATHINAM, *Moment growth bounds on continuous time Markov processes on non-negative integer lattices*, Quart. Appl. Math. 73 (2015), pp. 347–364.
- [20] M. RATHINAM, P. W. SHEPPARD, AND M. KHAMMASH, *Efficient computation of parameter sensitivities of discrete stochastic chemical reaction networks*, J. Chem. Phys., 132 (2010), 034103.
- [21] A. SALTELLI, P. ANNONI, I. AZZINI, F. CAMPOLONGO, M. RATTO, AND S. TARANTOLA, *Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index*, Comput. Phys. Commun., 181 (2010), pp. 259–270.
- [22] A. SALTELLI, M. RATTO, T. ANDRES, F. CAMPOLONGO, J. CARIBONI, D. GATELLI, M. SAISANA, AND S. TARANTOLA, *Global Sensitivity Analysis: The Primer*, John Wiley & Sons, Chichester, UK, 2008.
- [23] P. W. SHEPPARD, M. RATHINAM, AND M. KHAMMASH, *A pathwise derivative approach to the computation of parameter sensitivities in discrete stochastic chemical systems*, J. Chem. Phys., 136 (2012), 034115.
- [24] R. C. SMITH, *Uncertainty Quantification: Theory, Implementation, and Applications*, Comput. Sci. Engrg. 12, SIAM, Philadelphia, 2013.
- [25] I. M. SOBOLOV, *Sensitivity estimates for nonlinear mathematical models*, Math. Modeling Comput. Experiment, 1 (1993), pp. 407–414.
- [26] I. M. SOBOLOV, *Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates*, Math. Comput. Simulation, 55 (2001), pp. 271–280.
- [27] J. M. VILAR, H. Y. KUEH, N. BARKAI, AND S. LEIBLER, *Mechanisms of noise-resistance in genetic oscillators*, Proc. Natl. Acad. Sci. USA, 99 (2002), pp. 5988–5992.
- [28] T. WANG, *Parametric Sensitivity Analysis of Stochastic Reaction Networks*, Ph.D. thesis, University of Maryland, Baltimore County, Baltimore, MD, 2015.
- [29] T. WANG AND M. RATHINAM, *Efficiency of the Girsanov transformation approach for parametric sensitivity analysis of stochastic chemical kinetics*, SIAM/ASA J. Uncertain. Quantif., 4 (2016), pp. 1288–1322, <https://doi.org/10.1137/140998111>.
- [30] D. WILKINSON, *Stochastic Modelling for Systems Biology*, 2nd ed., CRC Press, Boca Raton, FL, 2012.