PROBABILISTIC SENSITIVITY ANALYSIS OF BIOCHEMICAL REACTION SYSTEMS

by

Hongxuan Zhang

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

October, 2010

Abstract

Sensitivity analysis is an indispensable tool for studying the robustness and fragility properties of biochemical reaction systems. In this thesis, we develop a probabilistic approach to sensitivity analysis, by combining a thermodynamically consistent probabilistic model of parameter fluctuations with an attractive decomposition scheme for the response variance. This approach addresses many problems associated with extensively used derivative-based sensitivity analysis techniques. Most importantly, it produces thermodynamically consistent sensitivity analysis results, allows different input factors to simultaneously fluctuate within a wide range of the parameter space, and can be effectively used to globally identify biochemical factors that influence selected system responses.

Variance-based sensitivity analysis requires evaluation of indices that cannot be done analytically. These are usually estimated by Monte-Carlo simulation, which is computationally demanding. Motivated by this problem, we study four approximation techniques that can be used to approximate the variance-based sensitivity indices. We highlight important theoretical, numerical, and computational aspects of each method, in an attempt to provide a comprehensive understanding of the advantages and disadvantages of each technique. It turns out that the computational cost of these techniques are orders of magnitude smaller than that of Monte Carlo estimation.

Experimental uncertainty about the true values of the kinetic parameters can greatly affect the accuracy of sensitivity analysis results. It is therefore important to develop a technique that minimizes the effects of this uncertainty. In this thesis, we extend our previous probabilistic model for the kinetic parameters to account for both biological and experimental variability and propose a new set of noise-reduced variance-based sensitivity indices. These indices are most suitable for sensitivity analysis of biochemical systems with poorly determined kinetic parameter values. We study three numerical techniques that can be used to evaluate the new sensitivity indices with appreciable computational savings.

By employing a computational model of mitogen-activated protein kinase signaling cascade, we demonstrate that our approach is well-suited for sensitivity analysis of biochemical reaction systems and can produce a wealth of information about the sensitivity properties of such systems.

Advisor: Dr. John I. Goutsias

Second Reader: Dr. Pablo A. Iglesias

Acknowledgments

First, I would like to thank Dr. John Goutsias for all his guidance since August 2007. Without his aid, I would not have been able to accomplish so much during my time at Johns Hopkins. Working with him has given me a more rigorous attitude in engineering research and has helped me retain and strengthen my intellectual curiosity. I appreciate his guidance and wisdom concerning my future aspirations as a researcher.

In addition, I would like to thank Dr. Pablo Iglesias for helping me understand the fundamentals of systems biology. I also thank him for taking time to read my dissertation and appreciate all his advices about my research. I would also like to thank Dr. Trac Tran for his willingness to serve in my GBO, seminar, and defense committees. I am glad that I had the opportunity to take two of his courses at Johns Hopkins.

I also want to express my gratitude to my lab mates, especially Garrett Jenkinson. Our discussions have been very stimulating and helpful. I appreciate his effort and willingness to help me whenever he could. In addition, I would like to acknowledge the National Science Foundation for financially supporting my research.

Finally, I would like to thank my wife, Hong, for helping me in so many ways and for making my life in Baltimore much more delightful. Meeting her is a special blessing

during my time at Johns Hopkins. I appreciate my parents for supporting me at both good and bad times. Without their love, I would not have been able to finish this PhD study.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Biological systems are more than simply a collection of molecules, cells, or organs. To precisely describe, model, and simulate biological systems, we need to understand how their individual components dynamically interact with each other to initiate, maintain, or modify biological activities responsible for cellular function and fate. This area of research is known as systems biology and requires effective integration of experimental and computational tools [1, 2].

A fundamental problem in computational systems biology is the construction of biochemical reaction system models that can effectively predict cellular behavior [1, 2]. Subsequent analysis of such models may reveal a wealth of biologically relevant information, including a list of biochemical factors (e.g., biochemical reactions and molecular species) that are most influential in shaping cellular responses. Determining the most influential factors in a biochemical reaction system is an important problem with many applications. For example, identifying influential biochemical factors and targeting these factors with high specificity is a promising pharmacological intervention

approach for treating human diseases [3].

A powerful tool for studying the properties of a biochemical reaction system is *sensitivity analysis* [4, 5]. The objective of sensitivity analysis is to determine the biochemical factors that produce no noticeable variations in system response and identify those factors that are most influential in shaping that response. Sensitivity analysis has been applied in many diverse fields of science and engineering, including mechanical engineering [6, 7], environmental engineering [8, 9], pharmacology [3, 10], biochemistry [5, 11, 12], and finance [13–15]. In systems biology, sensitivity analysis has allowed researchers to identify factors controlling biological behavior in cells [16–19], simplify procedures for designing and optimizing genetic circuits [20], obtain insights into the robustness and fragility tradeoff in cell regulation [21], and determine appropriate targets for pharmacological intervention [10, 22].

The sensitivity analysis approaches available in the literature can be generally classified into two groups. The first group deals primarily with deterministic techniques for sensitivity analysis, primarily based on derivatives of a response function of interest with respect to system parameters. The second group deals with probabilistic techniques, which quantify statistical variations in system response due to random perturbations in factors of interest.

Derivative-based sensitivity analysis techniques are subject to several drawbacks, which must be carefully considered before applying these techniques to problems in systems biology. First of all, derivative-based sensitivity analysis is limited to evaluating the effects of *infinitesimal* changes in parameter values on the system response. It will most certainly fail to reveal important sensitivity properties due to appreciable

parameter variations. In certain cases, we may be able to address this problem by averaging derivative values calculated at appropriately selected points in the parameter space [11]. It is expected however that, for appreciable parameter variations, such averaging will not provide an accurate description of the sensitivity properties of a biochemical reaction system. As a matter of fact, it has been pointed out by Saltelli *et al.* [23] that, although derivative averaging leads to a useful method for sensitivity analysis (known as *elementary effect test*), it should only be used to derive approximate sensitivity information, since it may introduce appreciable errors in the analysis.

In addition to the above, the task of accurately calculating response derivatives is not straightforward. One may easily express the response derivatives in terms of concentration sensitivities and analytically derive a system of differential equations that govern the dynamic evolution of these sensitivities. Then, evaluation of response derivatives will require simultaneous integration of the sensitivity equations together with the differential equations governing the dynamic evolution of the underlying molecular concentrations. Most often, this step cannot be implemented in a reasonable time, due to stiffness of the underlying differential equations [4]. As a consequence, most users of derivative-based sensitivity analysis techniques resort to approximating derivatives by finite-differences. However, the resulting approximations must be carefully used in applications, since it is difficult to theoretically predict, control, and numerically evaluate their accuracy [4, 11].

Finally, most derivative-based sensitivity analysis techniques in the literature use first-order derivatives to quantify the influence of a *single* parameter on the system response, while fixing the remaining parameters to their reference values. This is not

an appropriate strategy in most problems of systems biology, since control of biological behavior is usually exerted by the orchestrated influence of many biochemical factors. In our opinion, sensitivity analysis of a biochemical reaction system should consider the *simultaneous* influence of various biochemical factors on the system response. This is of particular interest in pharmacological applications, since system-based drug design techniques often consider the biological effects of targeting several biochemical factors simultaneously [24]. Although sensitivities with respect to two or more factors can be well-defined by means of second- and higher-order derivatives, accurate evaluation of such derivatives is a much more difficult problem than calculating first-order derivatives. Most importantly, due to their local nature, these derivatives cannot capture real joint sensitivity effects, which often occur at appreciable levels of parameter variations.

Due to the above drawbacks, we believe that the general use of derivative-based sensitivity analysis techniques in systems biology should be limited, despite their extensive use in the literature. In this dissertation we demonstrate that probabilistic sensitivity analysis, and more precisely the variance-based sensitivity analysis approach developed by Sobol', Saltelli, and their collaborators [23, 25–27], is better suited for biochemical reaction system analysis. Variance-based sensitivity analysis can effectively address the previous drawbacks associated with derivative-based techniques. It can easily accommodate appreciable parameter variations, and allows for a systematic investigation of interactions among different system components. Besides, variance-based sensitivity analysis is independent from the additivity or linearity of the system model, and is able to treat grouped factors as if they were single ones. The price to be paid however is a substantial increase in computational complexity, due to the high dimensional

integrations required for the evaluation of variance-based sensitivity indices. This problem must be addressed by devising sufficiently accurate approximation techniques to estimate the sensitivity indices efficiently.

Other probabilistic sensitivity analysis approaches have been introduced in the literature, such as a technique based on entropy and mutual information [7, 28]. Unfortunately, entropy-based sensitivity analysis requires using histograms to approximate continuous probability density functions, which is biased and prone to discretization errors, thus making estimation of factor interactions very difficult. Although corrections can be made to the biased entropy estimators, these corrections may not always be satisfactory [29]. Besides, due to the increased computational complexity associated with this approach, it is very hard to devise accurate approximation techniques to estimate entropy-based sensitivity indices efficiently. As a consequence, this approach is impractical for large biochemical reaction systems.

A special requirement in sensitivity analysis of biochemical reaction systems is to satisfy a number of necessary thermodynamic constraints, which may strongly limit the space of valid kinetic parameter values. Unfortunately, traditional sensitivity analysis approaches for biochemical reaction systems often ignore these important constraints, and may generate misleading results since physically impossible parameter values are also considered in the formulation [30]. Motivated by this issue, we propose in this dissertation a biophysically derived probabilistic model for reaction rate constants, so that randomly perturbed input factors can automatically satisfy the required thermodynamic constraints.

Variance-based sensitivity indices are usually estimated by Monte Carlo simulation [23, 27, 31, 32], which requires evaluation of the system response at each sampled parameter set. A major drawback of this approach is its slow rate of convergence. As a matter of fact, the error produced by a naive Monte Carlo estimation approach decreases with an error rate of $O(1/\sqrt{L})$, where L is the number of Monte Carlo samples used [33]. Hence, accurate estimation of the sensitivity indices requires a large number of Monte Carlo samples and, therefore, a large number of system response evaluations. This makes Monte Carlo estimation of variance-based sensitivity indices computationally very expensive, especially in the case of biochemical reaction systems comprised of many reactions and molecular species.

To reduce the computational burden of Monte Carlo simulations, it is imperative that we develop techniques which can result in sufficiently accurate and efficiently implementable estimators of sensitivity indices. In this dissertation, we present four such techniques, derivative approximation (DA), polynomial approximation (PA), Gauss-Hermite integration (GHI), and orthonormal Hermite approximation (OHA), and apply them to a well-known biochemical reaction model of the mitogen-activated protein kinase (MAPK) signaling cascade. We use this model in this dissertation to compare and assess the discussed sensitivity analysis techniques. DA is based on a second-order Taylor series expansion of the response function and is an extension of the first-order derivative-based approach for variance-based sensitivity analysis discussed in [5, 23, 27, 34] by including second-order derivative terms. The other three approximation techniques are based on the high-dimensional model representation (HDMR) schemes developed by H. Rabitz and his coworkers [35–37]. We derive analytical for-

mulas for the sensitivity indices generated by these four techniques, which allows to estimate their values more accurately and efficiently.

As it is the case for derivative-based sensitivity analysis techniques, the application of variance-based sensitivity analysis approaches requires specifying the nominal values of kinetic parameters. Moreover, the sensitivity results may strongly depend on the particular choice for these values. For the sensitivity results to be biologically relevant, the nominal values must be the true values. However, the true parameter values of a real biochemical reaction system are rarely known. Instead, they are substituted by estimated values, whose accuracy is often affected by unpredictable experimental variability. The issue here is that different nominal parameter values may produce different sensitivity analysis results. In many systems-biology applications, such as system-based drug target selection [16–18, 22], the main objective of sensitivity analysis is to assess how biological variability influences cellular behavior, in which case a great level of experimental uncertainty in the nominal values of kinetic parameters can significantly affect the variance-based sensitivity analysis results. Motivated by this problem, we propose in this dissertation a set of noise-reduced variance-based sensitivity indices, which can be used to exploit the effects of biological variability by appropriately averaging experimental variability out of the problem. We derive the proposed indices by extending the definitions of traditional variance-based sensitivity indices. It turns out that the new indices are more robust to the choice of nominal kinetic parameter values, can accommodate different levels of experimental uncertainty associated with different tools used to generate systems-level data [38–42], and lead to a powerful sensitivity analysis approach under experimental uncertainty.

This dissertation is structured as follows.

In Chapter 2, we provide a mathematical description for biochemical reaction systems, modeled by classical chemical kinetics. Moreover, we identify appropriate system response functions for sensitivity analysis based on a previously suggested model of the MAPK signaling cascade model. By using basic thermodynamic principles, we develop a probabilistic model for the rate constants of a biochemical reaction system. We employ this model to identify appropriate biochemical factors of interest to sensitivity analysis and mathematically characterize their fluctuations.

In Chapter 3, we present a variance-based sensitivity analysis technique and discuss its applicability to biochemical reaction systems. By limiting our interest to first- and second-order effects, we present a systematic methodology for classifying biochemical factors (reactions and molecular species) based on how these factors influence the system response. Then, we develop a set of Monte Carlo estimators for the quantities required by the proposed variance-based sensitivity analysis approach. By using the results obtained by applying our probabilistic sensitivity analysis approach on the MAPK signaling cascade, we identify the reactions and molecular species that are most important for controlling appropriately chosen response characteristics of this cascade. Our analysis agrees well with published experimental results and clearly demonstrates the potential of variance-based techniques for sensitivity analysis of biochemical reaction systems.

In Chapter 4, we develop four efficient methods that one can use to analytically approximate the variance-based sensitivity indices. We highlight important theoretical, numerical, and computational aspects of each method, in an attempt to provide a

comprehensive understanding of their advantages and disadvantages. In order to clarify the relative merits of each approximation technique and produce useful insights on when these techniques can be used for sensitivity analysis, we compare the results obtained by these methods with the results obtained by Monte Carlo estimation,

In Chapter 5, we extend the thermodynamically consistent probabilistic model for the reaction rate constants in order to account for uncertainty in their nominal values. This allows us to mathematically characterize fluctuations in biochemical factor values under both biological and experimental variability. We propose a set of noise-reduced variance-based sensitivity indices, which are designed to quantify the average relative importance of each biochemical factor under random perturbations in nominal kinetic parameter values. Experimental variability is accounted for and separated from biological variability, so that the results of sensitivity analysis are robust to variations in nominal parameter values. We use a well-known variance decomposition formula to derive Monte Carlo estimators for evaluating these noise-reduced variance-based sensitivity indices by avoiding direct averaging. This produces substantial improvement of numerical efficiency and stability. We also discuss two other numerical methods for estimating the proposed noise-reduced variance-based sensitivity indices more efficiently, using derivative approximation and dimensionality reduction based on orthonormal Hermite approximation. Again, we use the MAPK signaling cascade model to demonstrate various aspects of the sensitivity analysis approach using the new sensitivity indices.

Finally, we conclude the dissertation in Chapter 6, where we reiterate our findings and suggest future research.

Chapter 2

Biochemical Reaction Systems

2.1 Chemical Kinetics

In this dissertation, we consider a well-stirred (homogeneous) biochemical reaction system at constant temperature and volume that consists of M coupled *reversible* reactions:

$$\sum_{n=1}^{N} \nu_{nm} X_n \overset{k_{2m-1}}{\underset{k_{2m}}{\rightleftharpoons}} \sum_{n=1}^{N} \nu'_{nm} X_n, \qquad m = 1, 2, \ldots, M, \tag{2.1}$$

where k_{2m-1}, k_{2m} are the rate constants of the forward and reverse reactions and $\nu_{nm}, \nu'_{nm} \geq 0$ are the stoichiometry coefficients of the reactants and products. The system contains N molecular species $X_1, X_2, \ldots, X_N$ whose concentrations at time $t \geq 0$ are denoted by $x_1(t), x_2(t), \ldots, x_N(t)$, respectively. Here, we set the unit of concentration to be molecules/cell so that the units of the rate constants for monomolecular and multimolecular reactions can be normalized to s^{-1}. If we assume that the molecular concentrations evolve continuously as a function of time and that all reactions are sufficiently characterized by the mass action rate law, then we can characterize the dynamic

evolution of molecular concentrations by the following chemical kinetic equations [43]:

$$\frac{dx_n(t)}{dt} = \sum_{m=1}^{M} s_{nm} \left\{ k_{2m-1} \prod_{i=1}^{N} \left[x_i(t)\right]^{\nu_{im}} - k_{2m} \prod_{i=1}^{N} \left[x_i(t)\right]^{\nu'_{im}} \right\}, \quad t \geq 0, \;\; n = 1, 2, \ldots, N, \tag{2.2}$$

where

$$s_{nm} := \nu'_{nm} - \nu_{nm} \tag{2.3}$$

is the *net* stoichiometry coefficient of the n^{th} molecular species associated with the m^{th} reaction.

2.2 Example: MAPK Signaling Cascade

To illustrate various aspects of the sensitivity analysis techniques discussed in this dissertation, we focus on a specific biochemical reaction system that models the well-known mitogen-activated protein kinase (MAPK) signaling cascade. The MAPK signaling cascade is an important signaling pathway that couples the binding of growth factors to cell surface receptors with intracellular responses that control cellular growth, proliferation, differentiation, and survival [44]. We use a rather detailed model of this pathway, depicted in Fig. 2.1, introduced in the literature by Schoeberl *et al.* [45]. This model consists of $N = 23$ molecular species that interact with each other through $M = 21$ reactions, of which 11 reactions are reversible, whereas, the remaining 10 reactions are irreversible. For an irreversible reaction, we set the rate constant of the corresponding reverse reaction equal to zero. For simplicity, we have removed all reactions that characterize signal transduction from the epidermal growth factor (EGF) receptor to Ras-GTP, which is the input to the MAPK signaling cascade.

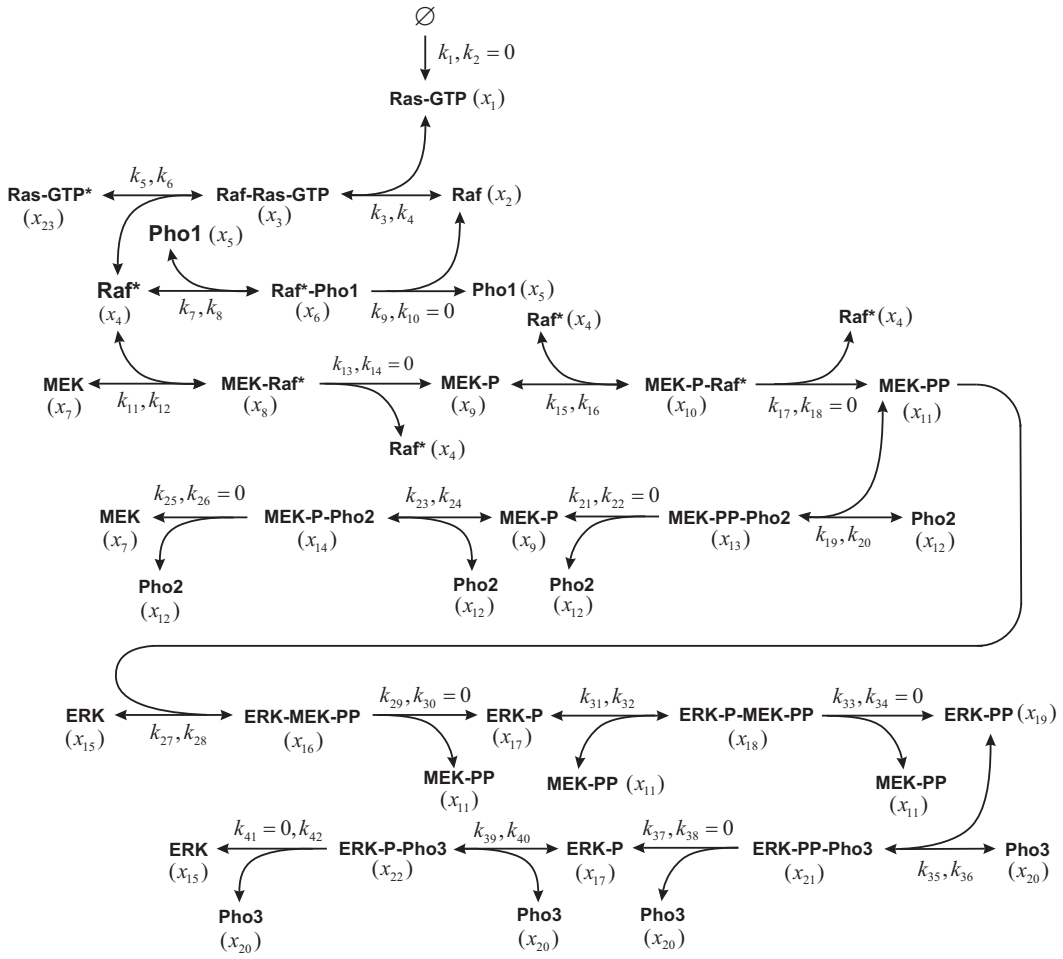

Figure 2.1: A biochemical reaction model of the MAPK signaling cascade.

The MAPK model begins with the synthesis of Ras-GTP, which interacts with Raf kinase to produce an active version Raf* of Raf; see Fig. 2.1. Raf* is capable of producing a doubly phosphorylated active version MEK-PP of the kinase MEK by two successive phosphorylation steps, whereas, MEK-PP can produce a doubly phosphorylated active version ERK-PP of the extracellular signal-regulated kinase (ERK) by two phosphorylation events as well. Each phosphorylation step is considered to be irreversible, unless mediated by an inactivating phosphatase. In this case, Pho1, Pho2, and Pho3 model inactivating phosphatases for Raf, MEK, and ERK, respectively.

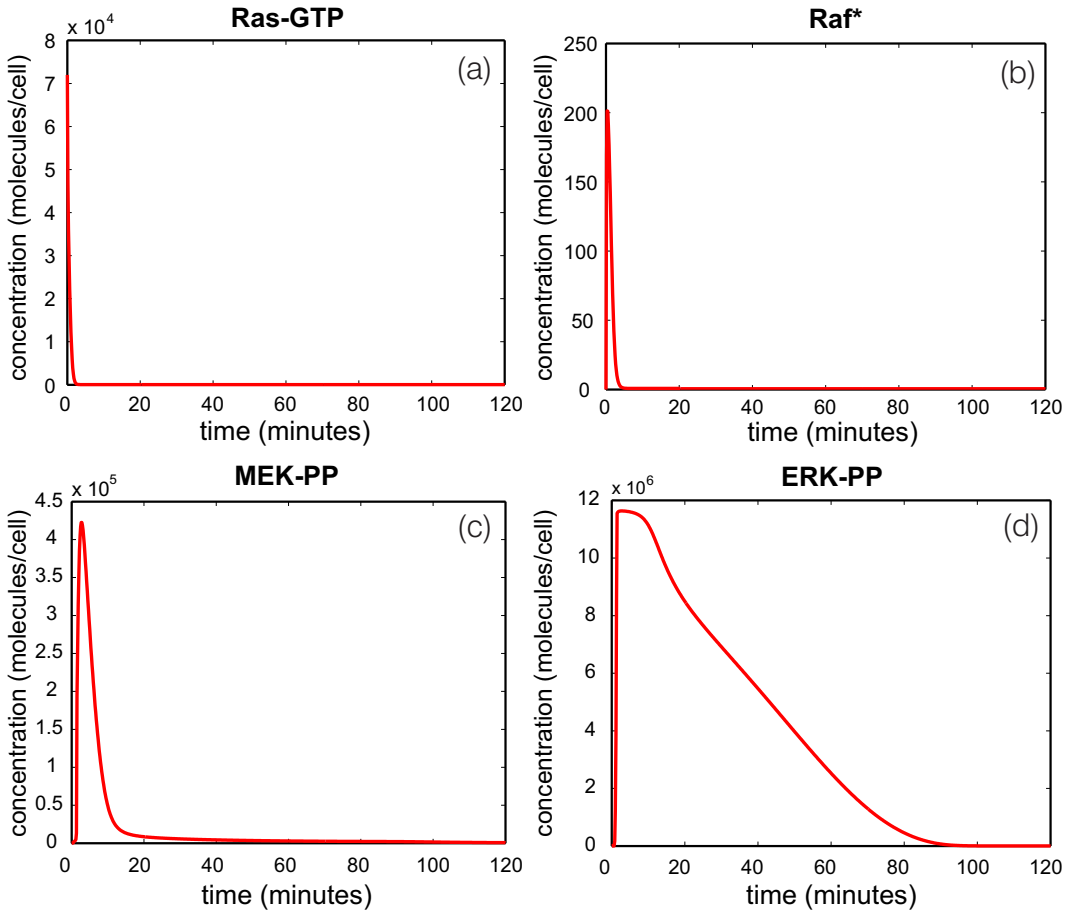

Figure 2.2: Concentration profiles of: (a) Ras-GTP, (b) Raf*, (c) MEK-PP, and (d) ERK-PP, predicted by the MAPK signaling cascade model depicted in Fig. 2.1.

In Tables 2.1 and 2.2, we list the biochemical reactions associated with the MAPK signaling cascade model and provide nominal values for the reaction rate constants. For irreversible reaction, we set the rate of the corresponding reverse reaction equal to zero. In Table 2.3, we provide values for the initial concentrations of the molecular species involved. These data are adopted from Schoeberl *et al.* [45], with a few rate constant values updated from the "JWS Online Cellular Systems Modeling" web site (http://jjj.biochem.sun.ac.za). The first reaction in the model depicted in Fig. 2.1 compensates for Ras-GTP synthesis, which, in reality, is accomplished by a complex epidermal growth factor (EGF)-induced signaling pathway [45]. We have set the reac-

Table 2.1: Reactions in the MAPK model and nominal values of the rate constants.

No.	Reaction	Rate Constant (s^{-1})
1	$\emptyset \rightarrow$ Ras-GTP	$k_1 = 3$
	Ras-GTP $\rightarrow \emptyset$	$k_2 = 0$
2	Ras-GTP + Raf $\rightarrow$ Raf-Ras-GTP	$k_3 = 1.66 \times 10^{-6}$
	Raf-Ras-GTP $\rightarrow$ Raf + Ras-GTP	$k_4 = 5.3 \times 10^{-3}$
3	Raf-Ras-GTP $\rightarrow$ Raf* + Ras-GTP*	$k_5 = 1$
	Raf* + Ras-GTP* $\rightarrow$ Raf-Ras-GTP	$k_6 = 1.16 \times 10^{-6}$
4	Raf* + Pho1 $\rightarrow$ Raf*-Pho1	$k_7 = 1.18 \times 10^{-4}$
	Raf*-Pho1 $\rightarrow$ Raf* + Pho1	$k_8 = 0.2$
5	Raf*-Pho1 $\rightarrow$ Raf + Pho1	$k_9 = 1$
	Raf + Pho1 $\rightarrow$ Raf*-Pho1	$k_{10} = 0$
6	MEK + Raf* $\rightarrow$ MEK-Raf*	$k_{11} = 1.94 \times 10^{-5}$
	MEK-Raf* $\rightarrow$ MEK + Raf*	$k_{12} = 3.3 \times 10^{-2}$
7	MEK-Raf* $\rightarrow$ MEK-P + Raf*	$k_{13} = 3.5$
	MEK-P + Raf* $\rightarrow$ MEK-Raf*	$k_{14} = 0$
8	MEK-P + Raf* $\rightarrow$ MEK-P-Raf*	$k_{15} = 1.94 \times 10^{-5}$
	MEK-P-Raf* $\rightarrow$ MEK-P + Raf*	$k_{16} = 3.3 \times 10^{-2}$
9	MEK-P-Raf* $\rightarrow$ MEK-PP + Raf*	$k_{17} = 2.9$
	MEK-PP + Raf* $\rightarrow$ MEK-P-Raf*	$k_{18} = 0$
10	MEK-PP + Pho2 $\rightarrow$ MEK-PP-Pho2	$k_{19} = 2.37 \times 10^{-5}$
	MEK-PP-Pho2 $\rightarrow$ MEK-PP + Pho2	$k_{20} = 0.8$

tion rate constant of Ras-GTP synthesis equal to $3s^{-1}$. This value results in an ERK-PP concentration profile that is similar to the one reported by Schoeberl *et al.* [45], with 50ng/ml EGF; compare Fig. 2.2(d) with Fig. 2(F) in Schoeberl *et al.* [45].

Using the given nominal reaction rate constant values and the initial molecular concentrations, we calculate the concentration profiles of all molecular species in the

Table 2.2: Reactions in the MAPK model and nominal values of the rate constants (cont.).

No.	Reaction	Rate Constant (s^{-1})
11	MEK-PP-Pho2 $\rightarrow$ MEK-P + Pho2	$k_{21} = 5.8 \times 10^{-2}$
	MEK-P + Pho2 $\rightarrow$ MEK-PP-Pho2	$k_{22} = 0$
12	MEK-P + Pho2 $\rightarrow$ MEK-P-Pho2	$k_{23} = 4.48 \times 10^{-7}$
	MEK-P-Pho2 $\rightarrow$ MEK-P + Pho2	$k_{24} = 0.5$
13	MEK-P-Pho2 $\rightarrow$ MEK + Pho2	$k_{25} = 5.8 \times 10^{-2}$
	MEK + Pho2 $\rightarrow$ MEK-P-Pho2	$k_{26} = 0$
14	ERK + MEK-PP $\rightarrow$ ERK-MEK-PP	$k_{27} = 8.87 \times 10^{-5}$
	ERK-MEK-PP $\rightarrow$ ERK + MEK-PP	$k_{28} = 1.83 \times 10^{-2}$
15	ERK-MEK-PP $\rightarrow$ ERK-P + MEK-PP	$k_{29} = 16$
	ERK-P + MEK-PP $\rightarrow$ ERK-MEK-PP	$k_{30} = 0$
16	ERK-P + MEK-PP $\rightarrow$ ERK-P-MEK-PP	$k_{31} = 8.87 \times 10^{-5}$
	ERK-P-MEK-PP $\rightarrow$ ERK-P + MEK-PP	$k_{32} = 1.83 \times 10^{-2}$
17	ERK-P-MEK-PP $\rightarrow$ ERK-PP + MEK-PP	$k_{33} = 5.7$
	ERK-PP + MEK-PP $\rightarrow$ ERK-P-MEK-PP	$k_{34} = 0$
18	ERK-PP + Pho3 $\rightarrow$ ERK-PP-Pho3	$k_{35} = 2.34 \times 10^{-5}$
	ERK-PP-Pho3 $\rightarrow$ ERK-PP + Pho3	$k_{36} = 0.6$
19	ERK-PP-Pho3 $\rightarrow$ ERK-P + Pho3	$k_{37} = 0.25$
	ERK-P + Pho3 $\rightarrow$ ERK-PP-Pho3	$k_{38} = 0$
20	ERK-P + Pho3 $\rightarrow$ ERK-P-Pho3	$k_{39} = 8.30 \times 10^{-6}$
	ERK-P-Pho3 $\rightarrow$ ERK-P + Pho3	$k_{40} = 0.5$
21	ERK + Pho3 $\rightarrow$ ERK-P-Pho3	$k_{41} = 0$
	ERK-P-Pho3 $\rightarrow$ ERK + Pho3	$k_{42} = 0.25$

MAPK signaling cascade. We depict four of these concentrations in Fig. 2.2.

Table 2.3: Initial molecular concentrations in the MAPK model.

No.	species	molecules/cell
1	Ras-GTP	7.20×10^4
2	Raf	4.00×10^4
3	Raf-Ras-GTP	0
4	Raf*	0
5	Pho1	4.00×10^4
6	Raf*-Pho1	0
7	MEK	2.10×10^8
8	MEK-Raf*	0
9	MEK-P	0
10	MEK-P-Raf*	0
11	MEK-PP	0
12	Pho2	4.00×10^4
13	MEK-PP-Pho2	0
14	MEK-P-Pho2	0
15	ERK	2.21×10^7
16	ERK-MEK-PP	0
17	ERK-P	0
18	ERK-P-MEK-PP	0
19	ERK-PP	0
20	Pho3	1.00×10^7
21	ERK-PP-Pho3	0
22	ERK-P-Pho3	0
23	Ras-GTP*	0

2.3 System Response Characteristics

Our objective is to use sensitivity analysis to quantify the relative importance of each reaction or molecular species in influencing a response characteristic of a biochemical reaction system. In general, different response characteristics will lead to different

sensitivity analysis results. Therefore, choosing an appropriate response characteristic is a very important issue. The response characteristic should not be chosen arbitrarily but through careful consideration of the biological problem at hand. However, any computable quantity that is thought to be important in influencing cellular function can serve as a useful response characteristic. This could be the entire time-varying concentration profile of a particular molecular species of interest (which leads to a time-dependent sensitivity analysis approach similar to the one suggested by Leloup and Goldbeter [21]), numerical features extracted from the profile, such as steady-state concentration or some quantities measuring the dynamical features of a concentration profile, or even numerical characteristics extracted from the flux of a selected reaction.

Here, we employ the MAPK signaling cascade as an example to demonstrate how to determine appropriate system responses. In MAPK, the doubly phosphorylated extracellular signal-regulated kinase (ERK-PP) is generally considered as the output molecular species. This species enters the nucleus of a cell and regulates translation of mRNA to proteins [46], as well as activities of transcription factors [47]. Therefore, and in order to find appropriate system responses, we focus our attention on the concentration profile of ERK-PP within an observation time interval $[0, t_{\max}]$.

It has been demonstrated in the literature that differences in the duration and strength of ERK-PP activity, produced by the MAPK signaling cascade, may generate distinct biological outcomes, such as cell differentiation, proliferation, and apoptosis [47–51]. Experimental evidence suggests that immediate early gene (IEG) products function as sensors for ERK-PP signal duration and strength [47, 49]. Moreover, it has been experimentally demonstrated that the time-integrated ERK-PP response directly

correlates with DNA synthesis [52, 53].

A conceptually simple way to explain these results is to assume the presence of a relatively unstable IEG product C, which is phosphorylated by ERK-PP resulting in stable molecules C-P. If we use the reaction

$$\text{C} + \text{ERK-PP} \xrightarrow{k} \text{C-P} + \text{ERK-PP}$$

to model phosphorylation of C, then the concentration profile $x_{\text{C-P}}(t)$ of C-P will satisfy the following differential equation:

$$\frac{dx_{\text{C-P}}(t)}{dt} = kx_{\text{C}}(t)x_{\text{ERK-PP}}(t). \tag{2.4}$$

As a consequence,

$$x_{\text{C-P}}(t) = kc \int_0^t x_{\text{ERK-PP}}(\tau)\, d\tau, \qquad \text{for} \quad t \geq 0, \tag{2.5}$$

where we assume for simplicity that the concentration of C remains constant for every $t \geq 0$. This shows that, at time t, the concentration of the stable phosphorylated product C-P will be proportional to the cumulative concentration of ERK-PP within the time interval $[0, t]$, in agreement with previously published results [53].

The activity of C-P can induce distinct biological outcomes by influencing transcriptional control. As a consequence, the integrated ERK-PP response is an important signaling characteristic for sensitivity analysis. It has been observed however that, due to certain biochemical factors (such as degradation and nuclear translocation), the integrated response of C-P may not be the only factor affecting cellular response. As a matter of fact, experimental evidence suggests that different biological outcomes may be produced by an activated MAP kinase, such as ERK-PP, depending on whether its

concentration remains above a critical level for a sufficient period of time [48, 52, 54]. Therefore, the duration and strength of ERK-PP activity are two additional signaling characteristics of importance to sensitivity analysis.

Based on the above discussion, we consider the following three response characteristics, namely the *duration* D, *integrated response* I, and *strength* S of ERK-PP, defined by

$$\begin{aligned} D &:= t_0, \\ I &:= \int_0^{t_0} x(t)dt, \\ S &:= \frac{1}{t_0}\int_0^{t_0} x(t)dt, \end{aligned} \tag{2.6}$$

where $x(t)$ is the concentration profile of ERK-PP and t_0 is the time at which $x(t)$ converges to zero. If convergence to zero does not occur within the observation time interval $[0, t_{\max}]$, then we set $t_0 = t_{\max}$. Here, strength is the average concentration during the time interval $[0, t_0]$. A practical way to approximate the duration is to determine the value of D such that

$$\int_0^D x(t)\, dt = (1-\varepsilon)\int_0^{t_{\max}} x(t)\, dt, \tag{2.7}$$

for a sufficiently small positive number ε.

We will take the system response to be the logarithm of the duration, integrated response, or strength. The reason for this choice is that, typically, the probability density functions of these quantities are long-tailed. Consequently, there is an appreciable probability for outliers, which may seriously compromise the numerical evaluation of response variances. Since we evaluate sensitivity indices in this dissertation by calculating conditional and unconditional response variances, we need to reduce the effect

of outliers. This objective can be achieved by considering log responses. We conjecture that a biological system is not sensitive to the response value per se but to its log value. Naturally, this provides robustness, with noticeable changes occurring only with strong and sustained variations in system response.

2.4 Probabilistic Modeling of Rate Constants

We can develop a powerful methodology for sensitivity analysis of biochemical reaction systems by recognizing that the rate constants of the underlying reactions may fluctuate randomly (e.g., due to unpredictable environmental, biological, and biochemical conditions) and by assessing how these fluctuations affect the system response. As it will become clear in the following, the forward and backward rate constants are not independent from each other. If these relationships among the system parameters are not taken into consideration, sensitivity analysis may lead to infeasible perturbed rate constants, which can easily break basic thermodynamic principles, such as detailed balance. This motivates us to express the rate constants by more fundamental parameters until a set of independent input factors can be obtained for sensitivity analysis. Our probabilistic model of rate constants is derived based on the classical Eyring theory, which is also called transition state theory or activated-complex theory [55].

Consider a simple binding reaction

$$X_1 + X_2 \overset{k_1}{\rightarrow} X_3. \tag{2.8}$$

According to Eyring theory, the reactants X_1 and X_2 form a hypothetical intermediate complex $(X_1X_2)^\ddagger$, known as the activated complex. This complex has a short lifetime and can be either transformed to the product X_3 or dissociate back to the reactants X_1 and X_2. We can describe these steps by the following reactions:

$$X_1 + X_2 \underset{k^\ddagger_{-1}}{\overset{k^\ddagger_1}{\rightleftharpoons}} (X_1X_2)^\ddagger \overset{k^\dagger_1}{\rightarrow} X_3. \tag{2.9}$$

A key assumption made in Eyring theory is that the reactants and the activated complex reach dynamic equilibrium at a time scale that is much shorter than the overall reaction time. In this case,

$$k^\ddagger_1 x_1(t)x_2(t) = k^\ddagger_{-1}(x_{12})^\ddagger(t), \qquad t \geq 0 \tag{2.10}$$

where $x_1(t)$, $x_2(t)$, and $(x_{12})^\ddagger(t)$ are the concentrations of X_1, X_2, and $(X_1X_2)^\ddagger$ respectively. The overall rate of product formation will be

$$\frac{dx_3(t)}{dt} = k^\dagger_1(x_{12})^\ddagger(t) = \frac{k_B T}{h}\frac{k^\ddagger_1}{k^\ddagger_{-1}}x_1(t)x_2(t), \tag{2.11}$$

where $k^\dagger_1 = k_B T/h$ is a universal constant for the activated complex given by statistical mechanics, with k_B being the Boltzmann constant ($k_B = 1.3806504 \times 10^{-23}\mathrm{JK}^{-1}$), T being the system temperature, and h being the Plank constant ($h = 6.62606885 \times 10^{-34}\mathrm{Js}$). In Eq. (2.11), the equilibrium constant can be further expressed by means of capacities [30] as

$$\frac{k^\ddagger_1}{k^\ddagger_{-1}} = \frac{c^\ddagger}{c_1 c_2}, \tag{2.12}$$

where c_1, c_2 are the capacities of the molecular species X_1 and X_2, respectively, and $c^\ddagger$ is the capacity of the activated complex $(X_1X_2)^\ddagger$. We can therefore express the overall

rate constant of the binding reaction in Eq. (2.8) as

$$k_1 = \frac{k_B T}{h} \frac{c^\ddagger}{c_1 c_2}. \tag{2.13}$$

Following a similar procedure for all forward and reverse reactions in the system, we can express their rate constants in terms of capacities as

$$\begin{aligned} k_{2m-1} &= \frac{k_B T}{h} \frac{c_m^\ddagger}{\prod_{n=1}^{N} c_n^{\nu_{nm}}}, \\ k_{2m} &= \frac{k_B T}{h} \frac{c_m^\ddagger}{\prod_{n=1}^{N} c_n^{\nu'_{nm}}}. \end{aligned} \tag{2.14}$$

These equations are known as Eyring-Polanyi equations [55]. The capacities are defined by [30]

$$\begin{aligned} c_m^\ddagger &:= x_{\text{total}} \exp\left\{-\frac{\mu_m^{\ddagger 0}}{k_B T}\right\}, \quad m = 1, 2, \ldots, M, \\ c_n &:= x_{\text{total}} \exp\left\{-\frac{\mu_n^{0}}{k_B T}\right\}, \quad n = 1, 2, \ldots, N, \end{aligned} \tag{2.15}$$

where x_{total} is the total concentration of all molecular species in the system that form a homogeneous molecular mixture in solution. This concentration is assumed to be constant, since for biochemical reaction systems, it normally contains a large fraction of inert species, such as water, whose concentration does not vary with time [30]. The quantity $\mu_m^{\ddagger 0}$ is the standard chemical potential of the activated complex associated with the m^{th} reaction, whereas, μ_n^0 is the standard chemical potential of the n^{th} molecular species. The chemical potential of a molecular species is the Gibbs free energy per mole of that species, and the standard chemical potential is the reference level of chemical potential under specified standard conditions. For ideal mixtures, the relationship between the chemical potential and the standard chemical potential of the n^{th} molecular

species in the system is given by

$$\mu_n(t) = \mu_n^0 + k_B T \ln \frac{x_n(t)}{x_{\text{total}}}, \tag{2.16}$$

where $x_n(t)$ is the concentration of the n^{th} molecular species.

Due to unpredictable biological variability, there is a great deal of uncertainty regarding the exact value of the standard chemical potential of a reaction. We may assume that the standard chemical potential of the activated complex associated with the m^{th} reaction is a random variable $M_m^{\ddagger 0}$, given by[1]

$$M_m^{\ddagger 0} = \mu_m^{\ddagger 0} + k_B T Y_m^{\ddagger}, \quad m = 1, 2, \ldots, M, \tag{2.17}$$

where $\mu_m^{\ddagger 0}$ is a nominal value and $Y_m^{\ddagger}$ is a zero mean random variable that accounts for variations of the standard chemical potential about its nominal value. A possible source of variation for the standard chemical potential of the activated complex is unpredictable fluctuations in the conformational states and concentrations of enzymes responsible for catalyzing the reaction [56, 57]. It is well known that enzymes enhance the reaction rate by lowering the standard chemical potential of the activated complex [58]. It is therefore expected that uncertainty in the conformational state and concentration of an enzyme catalyzing a given reaction will produce uncertainty in the value of the standard chemical potential of the activated complex of that reaction.

Similarly, we may assume that the standard chemical potential of the n^{th} molecular species is a random variable M_n^0, given by

$$M_n^0 = \mu_n^0 + k_B T Y_n, \quad n = 1, 2, \ldots, N, \tag{2.18}$$

[1] In this dissertation, we use capital letters to denote random variables and small letters to denote their realizations (samples).

where μ_n^0 is a nominal value and Y_n is a zero mean random variable that accounts for variations of the standard chemical potential about its nominal value. A possible source of variation in the standard chemical potential of a molecular species, such as a protein, is genetic point (single nucleotide) mutations, which may occur randomly and may result in a small change of the amino acid sequence of the protein. Because of redundancy of the genetic code (64 codons and only 20 amino acids) and the relative closeness within the code of biochemically similar amino acids, most small changes in proteins will not affect the stoichiometry of the underlying biochemical reaction network but may alter the kinetic properties of the participating molecular species by modifying the standard chemical potential (or capacity) values.

As a consequence of the previous discussion, we can now treat the rate constants as random variables K_{2m-1} and K_{2m}, given by

$$\begin{aligned} K_{2m-1} &= \frac{k_B T}{h} \frac{C_m^{\ddagger}}{\prod_{n=1}^{N} C_n^{\nu_{nm}}}, \\ K_{2m} &= \frac{k_B T}{h} \frac{C_m^{\ddagger}}{\prod_{n=1}^{N} C_n^{\nu'_{nm}}}, \end{aligned} \tag{2.19}$$

where the (random) capacities $C_m^{\ddagger}$ and C_n are defined by

$$\begin{aligned} C_m^{\ddagger} &:= x_{\text{total}} \exp\left\{-\frac{M_m^{\ddagger 0}}{k_B T}\right\}, \quad m = 1, 2, \ldots, M, \\ C_n &:= x_{\text{total}} \exp\left\{-\frac{M_n^0}{k_B T}\right\}, \quad n = 1, 2, \ldots, N. \end{aligned} \tag{2.20}$$

From Eqs. (2.14), (2.15), and (2.17)–(2.20), we have that

$$\begin{aligned} K_{2m-1} &= k_{2m-1} \exp\{-Y_m^{\ddagger}\} \exp\left\{\sum_{n=1}^{N} \nu_{nm} Y_n\right\}, \\ K_{2m} &= k_{2m} \exp\{-Y_m^{\ddagger}\} \exp\left\{\sum_{n=1}^{N} \nu'_{nm} Y_n\right\}, \end{aligned} \tag{2.21}$$

where k_{2m-1} and k_{2m} are given by Eqs. (2.14) and (2.15). Implicit in Eq. (2.21) is the assumption that, once the values of the random variables $\{Y_m^\ddagger,\, m = 1, 2, \ldots, M\}$ and $\{Y_n,\, n = 1, 2, \ldots, N\}$ are determined, these values remain constant during a time period of interest. It has been argued [59] that this is reasonable if we assume that the biochemical reaction system exhibits only static (and not dynamic) disorder.

In the following, we assume that $Y_m^\ddagger$ is a zero-mean Gaussian random variable with standard deviation $\lambda_m^\ddagger$; i.e., $Y_m^\ddagger \sim \mathcal{N}(0, \lambda_m^\ddagger)$, for $m = 1, 2, \ldots, M$. Then, Eq. (2.17) implies that the standard chemical potential $M_m^{\ddagger 0}$ of the activated complex of the m^{th} reaction is a Gaussian random variable with mean value $\mu_m^{\ddagger 0}$ and standard deviation $k_B T \lambda_m^\ddagger$. Similarly, we assume that Y_n is a zero-mean Gaussian random variable with standard deviation λ_n; i.e., $Y_n \sim \mathcal{N}(0, \lambda_n)$, for $n = 1, 2, \ldots, N$. In this case, Eq. (2.18) implies that the standard chemical potential M_n^0 of the n^{th} molecular species is a Gaussian random variable with mean value μ_n^0 and standard deviation $k_B T \lambda_n$. Finally, we assume that the random variables $\{Y_m^\ddagger, m = 1, 2, \ldots, M\}$ and $\{Y_n, n = 1, 2, \ldots, N\}$ are mutually independent.

The use of Gaussian distributions for modeling $Y_m^\ddagger$ and Y_n cannot be easily justified experimentally. However, it is a convenient choice that can be viewed as an effective approximation of the actual distributions of $Y_m^\ddagger$ and Y_n, obtained by setting all high-order (≥ 3) statistical moments equal to zero. As a consequence, the probability distributions of the forward and reverse reaction rates are log-normal. Interestingly, it has been argued in the literature that log-normal distributions are natural choices for modeling processes evolving by energy transduction mechanisms that reduce free energy [60], which is the case in biochemical reaction systems.

The assumption of statistical independence between the random variables $\{Y_m^\ddagger, m = 1, 2, \ldots, M\}$ and $\{Y_n, n = 1, 2, \ldots, N\}$ can be justified by arguing that the primary sources of variation in the standard chemical potentials of activated complexes and molecular species are different (e.g., fluctuations in enzyme concentrations vs. genetic point mutations) and do not influence each other. However, it is more difficult to justify mutual independence within $\{Y_m^\ddagger, m = 1, 2, \ldots, M\}$ and $\{Y_n, n = 1, 2, \ldots, N\}$. We simply view this as a convenient approximating assumption that allows us to proceed with the sensitivity analysis approach discussed in Chapter 3.

Eq. (2.21) suggests that variations in the forward and reverse reaction rates may occur due to variations in the standard chemical potential of the activated complex associated with that reaction and variations in the standard chemical potentials of its reactants. These variations are modeled by two Gaussian random variables

$$\begin{aligned} G_m &:= -Y_m^\ddagger + \textstyle\sum_{n=1}^N \nu_{nm} Y_n, \\ G'_m &:= -Y_m^\ddagger + \textstyle\sum_{n=1}^N \nu'_{nm} Y_n, \end{aligned} \tag{2.22}$$

for the forward and reverse reactions, respectively. Note that

$$\begin{aligned} &\mathrm{E}\big[G_m\big] = 0, \quad \mathrm{Var}\big[G_m\big] = \big[\lambda_m^\ddagger\big]^2 + \textstyle\sum_{n=1}^N \nu_{nm}^2 \lambda_n^2, \\ &\mathrm{E}\big[G'_m\big] = 0, \quad \mathrm{Var}\big[G'_m\big] = \big[\lambda_m^\ddagger\big]^2 + \textstyle\sum_{n=1}^N \big[\nu'_{nm}\big]^2 \lambda_n^2. \end{aligned} \tag{2.23}$$

Therefore, and due to the dependence of $\mathrm{Var}\big[G_m\big]$ and $\mathrm{Var}\big[G'_m\big]$ on the stoichiometry coefficients ν_{nm} and ν'_{nm}, the size of variations in the reaction rates will not be uniform in general, even if all standard deviations $\lambda_m^\ddagger$ and λ_n take the same values. Note also that the probability distributions of K_{2m-1} and K_{2m} are log-normal, with $\mathrm{median}\big[K_{2m-1}\big] = k_{2m-1}$ and $\mathrm{median}\big[K_{2m}\big] = k_{2m}$.

It has been recently pointed out in the literature that a serious issue associated with existing sensitivity analysis techniques is lack of thermodynamic consistency [30]. However, an important consequence of Eqs. (2.14) and (2.21) is that

$$\frac{k_{2m-1}}{k_{2m}} = \prod_{n=1}^{N} c_n^{s_{nm}} \qquad \text{and} \qquad \frac{K_{2m-1}}{K_{2m}} = \frac{k_{2m-1}}{k_{2m}} \exp\left\{ -\sum_{n=1}^{N} s_{nm} Y_n \right\}, \tag{2.24}$$

by virtue of Eq. (2.3), which are constraints thermodynamically imposed on the reaction rate constants. Note that, if $\mathbf{b}$ is an $M \times 1$ vector in the null space of the $N \times M$ stoichiometry matrix $\mathbb{S} = \left[s_{nm} \right]$ of the biochemical reaction system (i.e., if $\mathbb{S}\,\mathbf{b} = 0$), then Eq. (2.24) implies that

$$\prod_{m=1}^{M} \left(\frac{K_{2m-1}}{K_{2m}} \right)^{b_m} = \prod_{m=1}^{M} \left(\frac{k_{2m-1}}{k_{2m}} \right)^{b_m} = 1, \tag{2.25}$$

which are known as Wegscheider conditions or detailed balance relations [30, 43]. Eq. (2.25) shows that the proposed probabilistic model for the reaction rate constants automatically satisfies detailed balance, as long as these conditions are satisfied by the nominal (median) reaction rate constants k_m. As a consequence, and in sharp contrast to existing techniques, the sensitivity analysis approach we present in this dissertation can produce thermodynamically consistent results.

To use Eq. (2.21) for sensitivity analysis, we need to make sure that the nominal values of the reaction rate constants lead to a thermodynamically feasible biochemical reaction system, in the sense that there exist capacities $c_m^{\ddagger}$, $m = 1, 2, \ldots, M$, and c_n, $n = 1, 2, \ldots, N$, such that the Eyring-Polanyi equations are satisfied. Using the MAPK signaling cascade model described in the Section 2.2, we show here how to check whether the nominal rate constant values are thermodynamically feasible. In particular, we need to prove that there exist capacities $c_m^{\ddagger}$, $m = 1, 2, \ldots, M$, and c_n,

$n = 1, 2, \ldots, N$, so that Eq. (2.14) is satisfied. Note that

$$\frac{k_{2m-1}}{k_{2m}} = \prod_{n=1}^{N} c_n^{s_{nm}}, \quad \text{for every} \quad m \in \mathcal{M}_r, \tag{2.26}$$

from Eq. (2.24), where s_{nm} is given by Eq. (2.3) and $\mathcal{M}_r$ denotes the set of all reversible reactions in the MAPK signaling cascade (i.e., reactions 2–4, 6, 8, 10, 12, 14, 16, 18, and 20). If we denote by $\mathbb{S}_r$ the $N \times M_r$ stoichiometry matrix associated with the reversible reactions, by $\mathbf{c}$ the $N \times 1$ vector of the molecular capacities $c_n, n = 1, 2, \ldots, N$, and by $\mathbf{r}$ the $M_r \times 1$ vector of the reaction rate ratios (equilibrium constants) k_{2m-1}/k_{2m}, $m \in \mathcal{M}_r$, then we can write Eq. (2.26) above in the following matrix-vector form:

$$\mathbb{S}_r^T \ln \mathbf{c} = \ln \mathbf{r},$$

where $\ln \mathbf{u}$ denotes a vector with elements $\ln u_i$, where u_i is the i^{th} element of vector $\mathbf{u}$.

It turns out that, for the MAPK signaling cascade, the columns of the stoichiometry matrix $\mathbb{S}_r$ are linearly independent; i.e., $\text{rank}(\mathbb{S}_r) = M_r$. As a consequence, we can write

$$\mathbb{S}_r^T \ln \mathbf{c} = \begin{bmatrix} \mathbb{S}_* \\ \\ \mathbb{S}_{**} \end{bmatrix}^T \begin{bmatrix} \ln \mathbf{c}_1 \\ \\ \ln \mathbf{c}_2 \end{bmatrix} = \mathbb{S}_*^T \ln \mathbf{c}_1 + \mathbb{S}_{**}^T \ln \mathbf{c}_2 = \ln \mathbf{r}, \tag{2.27}$$

where $\mathbb{S}_*$ is the $M_r \times M_r$ matrix that contains all linearly independent rows of $\mathbb{S}_r$, $\mathbb{S}_{**}$ is the $(N - M_r) \times M_r$ matrix that contains the remaining rows of $\mathbb{S}_r$, and $\mathbf{c}_1$, $\mathbf{c}_2$ are the corresponding sub-vectors of $\mathbf{c}$. Note that $\mathbb{S}_*$ is an invertible matrix. Therefore, Eq. (2.27) above implies that

$$\ln \mathbf{c}_1 = (\mathbb{S}_*^T)^{-1} \left(\ln \mathbf{r} - \mathbb{S}_{**}^T \ln \mathbf{c}_2 \right). \tag{2.28}$$

We can now set $\ln \mathbf{c}_2 = 0$, in which case $\ln \mathbf{c}_1 = (\mathbb{S}_*^T)^{-1} \ln \mathbf{r}$. Therefore, given the reaction rate constants of the reversible reactions, we can find capacity values for all

molecular species in the system. Subsequently, we can determine the capacities of the activated complexes by setting

$$c_m^{\ddagger} = k_{2m-1} \frac{h}{k_B T} \prod_{n=1}^{N} c_n^{\nu_{nm}}, \quad \text{for} \;\; m = 1, 2, \ldots, M. \tag{2.29}$$

The previous discussion shows that the rate constant values we use in this dissertation correspond to a thermodynamically feasible model for the MAPK signaling cascade, since, given these values, we can find appropriate capacity values so that the Eyring-Polanyi equations are satisfied.

Finally, we should note that some reactions in Eq. (2.1) may be "incomplete," in the sense that they are expressed without indicating every molecular species participating in the reaction. For example, it is common to model phosphorylation of a protein A by a kinase K using the following reaction: $\mathrm{A} + \mathrm{K} \xrightarrow{k} \mathrm{A\text{-}P} + \mathrm{K}$. However, phosphorylation also includes the cleavage of adenosine triphosphate (ATP) into adenosine diphosphate (ADP) and inorganic phosphate (P), which leads to the following more accurate reaction: $\mathrm{A} + \mathrm{K} + \mathrm{ATP} \xrightarrow{k'} \mathrm{A\text{-}P} + \mathrm{K} + \mathrm{ADP}$. Note, however, that if we set $k = k' x_{\mathrm{ATP}}$, then the two reactions will be equivalent. Therefore, and in order to deal with an incomplete reaction, we will assume that its reaction rate is the actual rate multiplied by the corresponding concentrations of the "missing" reactants. In this case, fluctuations in the rate constant values may also be attributed to fluctuations in the concentrations of these reactants.

The issue of thermodynamic consistency has also been discussed by Liebermeister and Klipp [61], as well as by Schaber, Liebermeister and Klipp [62], who have proposed an alternative approach to the one presented in this dissertation for modeling the reaction rate constants of a biochemical reaction system. Their formulation is based

on the assumption that the geometric mean of the forward and reverse reaction rate constants is independent from the corresponding equilibrium constant, which may not be valid in general. For this reason, we chose to work here with the Eyring-Polanyi equations, which are derived by means of the well-known activated-complex theory, one of the most commonly used theory in chemical kinetics [55].

Chapter 3

Variance-Based Sensitivity Analysis

3.1 Definitions

In Section 2.4, we introduced a probabilistic model for the reaction rate constants of a biochemical reaction system derived by modeling random fluctuations in the standard chemical potentials of the activated complexes associated with the reactions and the underlying molecular species. By assessing how these fluctuations affect the system response, we can classify reactions (molecular species) into two groups, namely influential reactions (molecular species) and non-influential reactions (molecular species). We say that a reaction (molecular species) is influential if random fluctuations in the corresponding standard chemical potential result in noticeable fluctuations in the system response. Otherwise, the reaction (molecular species) is said to be non-influential.

In this chapter, we discuss a powerful approach to sensitivity analysis, known as variance-based sensitivity analysis [5, 23, 27]. To simplify notation, we will use $U_1, U_2, \ldots, U_J$ to denote the random variables $Y^{\ddagger}$ and Y associated with the stan-

dard chemical potentials. In general, we take $J = M + N$ and set $U_j = Y_j^{\ddagger}$, for $j = 1, 2, \ldots, M$, and $U_{M+j} = Y_j$, for $j = 1, 2, \ldots, N$. However, when the standard chemical potentials of the molecular species are fixed, we take $J = M$ and set $U_j = Y_j^{\ddagger}$, for $j = 1, 2, \ldots, M$, whereas, when the standard chemical potentials of the activated complexes are fixed, we take $J = N$ and set $U_j = Y_j$, for $j = 1, 2, \ldots, N$. We will be referring to $U_1, U_2, \ldots, U_J$ as "biochemical factors."

Given a system response function $R(\mathbf{u})$, we can easily verify that its variance satisfies the following equation:

$$V = \sum_{j=1}^{J} V_j + \sum_{j=1}^{J} \sum_{j'>j} V_{jj'} + \cdots + V_{12\cdots J}, \tag{3.1}$$

where

$$\begin{aligned} V &:= \mathrm{Var}\big[R(\mathbf{U})\big], \\ V_j &:= \mathrm{Var}\big[\mathrm{E}\big[R(\mathbf{U}) \mid U_j\big]\big], \\ V_{jj'} &:= \mathrm{Var}\big[\mathrm{E}\big[R(\mathbf{U}) \mid U_j, U_{j'}\big]\big] - \mathrm{Var}\big[\mathrm{E}\big[R(\mathbf{U}) \mid U_j\big]\big] - \mathrm{Var}\big[\mathrm{E}\big[R(\mathbf{U}) \mid U_{j'}\big]\big], \end{aligned} \tag{3.2}$$

with similar expressions for the remaining terms. If the biochemical factors $U_1, U_2, \ldots, U_J$ are statistically independent (which we have assumed here to be true), then each term on the right-hand-side of Eq. (3.1) will be nonnegative (the V_j terms are always nonnegative). This result was first shown by Sobol' [25, 26], and serves as the basis for constructing the sensitivity indices we discuss below.

Due to the well-known variance decomposition formula

$$\mathrm{Var}\big[Y\big] = \mathrm{Var}\big[\mathrm{E}\big[Y|X\big]\big] + \mathrm{E}\big[\mathrm{Var}\big[Y|X\big]\big], \tag{3.3}$$

Eq. (3.2) implies that

$$V_j = \mathrm{Var}\big[R(\mathbf{U})\big] - \mathrm{E}\big[\mathrm{Var}\big[R(\mathbf{U}) \mid U_j\big]\big]. \tag{3.4}$$

As a consequence, V_j quantifies the average reduction in response variance obtained by keeping the j^{th} biochemical factor fixed. Therefore, we may use V_j as a measure of the *singular* contribution of the j^{th} biochemical factor to the system response. Likewise, from Eqs. (3.2)–(3.4), we obtain

$$V_j + V_{j'} + V_{jj'} = \mathrm{Var}\big[R(\mathbf{U})\big] - \mathrm{E}\big[\mathrm{Var}\big[R(\mathbf{U}) \mid U_j, U_{j'}\big]\big]. \tag{3.5}$$

Clearly, the term $V_{jj'}$ quantifies the average reduction in the response variance due to jointly fixing the two biochemical factors U_j and $U_{j'}$, which is not accounted for by summing the average reductions obtained by separately fixing these factors. Hence, we may use $V_{jj'}$ as a measure of the *joint* contribution of the biochemical factors U_j and $U_{j'}$ to the system response. Similar remarks apply for the higher-order terms on the right-hand-side of Eq. (3.1). Clearly, Eq. (3.1) decomposes the response variance V of a biochemical reaction system into a sum of individual terms, where each term quantifies the singular or joint contribution of a particular biochemical factor to the system response.

If the response function $R(\mathbf{u})$ is sufficiently smooth around $\mathbf{0}$, so that its derivatives of order ≥ 3 at $\mathbf{0}$ are negligible, then its Taylor series expansion about $\mathbf{0}$ is given by

$$R(\mathbf{U}) \simeq R(\mathbf{0}) + \sum_{j=1}^{J} \frac{\partial R(\mathbf{0})}{\partial u_j}\, U_j + \frac{1}{2} \sum_{j=1}^{J} \sum_{j'=1}^{J} \frac{\partial^2 R(\mathbf{0})}{\partial u_j \partial u_{j'}}\, U_j U_{j'}. \tag{3.6}$$

For a zero-mean Gaussian random variable U_j with standard deviation λ_j, we have that

$$\mathrm{E}\big[U_j^3\big] = 0 \qquad \text{and} \qquad \mathrm{E}\big[U_j^4\big] = 3\lambda_j^4. \tag{3.7}$$

If the biochemical factors U_j have sufficiently small standard deviations λ_j, so that

$$\lambda_j^2 \lambda_{j'}^2 \left[\frac{\partial^2 R(\mathbf{0})}{\partial u_j \partial u_{j'}}\right]^2 \simeq\ 0, \quad \text{for every } \ j, j' = 1, 2, \ldots, J, \tag{3.8}$$

we obtain

$$\mathrm{E}\big[R(\mathbf{U})\big] \simeq R(\mathbf{0}), \tag{3.9}$$

and

$$V \simeq V_1 + V_2 + \cdots + V_J, \tag{3.10}$$

where

$$V_j \simeq \lambda_j^2 \left[\frac{\partial R(\mathbf{0})}{\partial u_j}\right]^2. \tag{3.11}$$

See also the discussion in [34]. In this case, we can approximately decompose the response variance into a sum of individual terms V_j, $j = 1, 2, \ldots, J$. Then, the biochemical factors U_j will mostly contribute to the response variance singularly. Note that, if $R(\mathbf{u})$ is *linear*, then Eqs. (3.10) and (3.11) will be exact, regardless of the statistical model assumed for the biochemical factors U_j. More generally, Eq. (3.10) (but not necessarily Eq. (3.11)) is exact when the response function is *additive*.[2] In both cases, it implies that all joint contributions of the biochemical factors $U_1, U_2, \ldots, U_J$ to the response variance in Eq. (3.1) will be zero.

In general, the total contribution of the j^{th} biochemical factor to the response variance is given by

$$C_j := V_j + \sum_{j' \neq j}^{J} V_{jj'} + \cdots + V_{12\cdots J}, \tag{3.12}$$

which implies that $V \leq \sum_{j=1}^{J} C_j$. Therefore, we cannot in general decompose the response variance into a sum of individual contributions from each biochemical factor.

From Eq. (3.12), we have that

$$\tau_j := \frac{C_j}{V} = \frac{V_j + \sum_{j' \neq j}^{J} V_{jj'} + \cdots + V_{12\cdots J}}{V}. \tag{3.13}$$

[2] A multivariate function is said to be *additive*, if it can be decomposed into a sum of one-dimensional functions of one variable.

This index has been introduced as a tool for sensitivity analysis by Saltelli and his collaborators [5, 23, 27, 63, 64]. It quantifies the fractional total (singular and joint) contribution of the j^{th} biochemical factor to the response variance. For this reason, we refer to τ_j as the *total-effect* sensitivity index (TESI) of the j^{th} biochemical factor. Note that

$$0 \leq \tau_j \leq 1, \quad \text{for every} \;\; j = 1, 2, \ldots, J, \tag{3.14}$$

due to Eqs. (3.1) and (3.14). Moreover, it can be shown that

$$\tau_j = \frac{\mathrm{E}\big[\mathrm{Var}\big[R(\mathbf{U}) \mid \mathbf{U}_{(j)}\big]\big]}{\mathrm{Var}\big[R(\mathbf{U})\big]} = 1 - \frac{\mathrm{Var}\big[\mathrm{E}\big[R(\mathbf{U}) \mid \mathbf{U}_{(j)}\big]\big]}{\mathrm{Var}\big[R(\mathbf{U})\big]}, \tag{3.15}$$

where $\mathbf{U}_{(j)}$ denotes the collection of all biochemical factors excluding U_j [23]. Therefore, τ_j is the average fractional response variance obtained when all factors, except the j^{th} biochemical factor, are kept fixed.

An important objective of sensitivity analysis is to identify *non-influential* biochemical factors (i.e., factors that do not appreciably influence the system response). This is of fundamental theoretical and experimental interest, since the biochemical reaction system will be *robust* to changes in the values of non-influential factors, which can be fixed without appreciably affecting the system response. As a consequence, we may be able to ignore non-influential biochemical factors, thus simplifying the complexity of the biochemical system under consideration. It is intuitive to believe that, if a biochemical factor is non-influential, fluctuations in its value will not appreciably affect the system response. This observation motivates us to define a biochemical factor as being *non-influential* if random fluctuations in its value does not generate noticeable fluctuations of the system response when the remaining factors are kept fixed. Based on this definition and Eq. (3.15), the j^{th} biochemical factor is non-influential if and

only if $\tau_j = 0$. Therefore, we can identify non-influential biochemical factors as those with TESI values close to zero.

After sorting out the non-influential biochemical factors, we would like to derive an index that we can use to rank the remaining factors based on how much these factors influence the system response. If we assume that Eq. (3.10) is satisfied, it will be natural to rank influential biochemical factors based on their corresponding V_j values, since V_j quantifies the contribution of the j^{th} biochemical factor to the response variance. This leads to ranking influential biochemical factors based on the following index:

$$\sigma_j := \frac{V_j}{V} = \frac{\mathrm{Var}\big[\mathrm{E}\big[R(\mathbf{U}) \mid U_j\big]\big]}{\mathrm{Var}\big[R(\mathbf{U})\big]}. \tag{3.16}$$

It is clear from this formula that σ_j quantifies the fractional singular contribution of the j^{th} biochemical factor to the response variance. Therefore, we will refer to this quantity as the *single-effect* sensitivity index (SESI) of the j^{th} biochemical factor. This index has been originally introduced by Iman [65], used by Krewski *et al.* [10], and then by Saltelli and his collaborators [5, 23, 27, 64]. Note that $0 \leq \sigma_j \leq 1$, for every $j = 1, 2, \ldots, J$, due to Eqs. (3.1) and (3.16). Thus, similar to the TESI's, the SESI's are also normalized to take values between 0 and 1.

Subject to the assumptions that lead to Eqs. (3.10) and (3.11), we have that

$$\tau_j \simeq \sigma_j \simeq \frac{\lambda_j^2 \left[\partial R(\overline{\mathbf{u}})/\partial u_j\right]^2}{\sum_{j'=1}^{J} \lambda_{j'}^2 \left[\partial R(\overline{\mathbf{u}})/\partial u_{j'}\right]^2}\ , \tag{3.17}$$

which provides a direct link between derivative-based and variance-based sensitivity analysis techniques and shows that derivative-based sensitivity analysis may be viewed as a special and restrictive case of variance-based sensitivity analysis. The form of SESI given by Eq. (3.17) has been proposed as a tool for sensitivity analysis by several

investigators [5, 23, 27, 34]. However, its use requires verification of the assumptions associated with Eqs. (3.10) and (3.11), which is clearly very difficult to do in practice.

From Eqs. (3.1), (3.13), and (3.16), note that $\sigma_j \leq \tau_j$, for every $j = 1, 2, \ldots, J$, whereas,

$$\sum_{j=1}^{J} \sigma_j \leq 1 \leq \sum_{j=1}^{J} \tau_j. \tag{3.18}$$

The difference

$$\delta := 1 - \sum_{j=1}^{J} \sigma_j \tag{3.19}$$

satisfies $0 \leq \delta \leq 1$. Moreover,

$$\delta = \frac{1}{V}\left(V - \sum_{j=1}^{J} V_j\right) = \frac{1}{V}\left(\sum_{j=1}^{J}\sum_{j'>j}^{J} V_{jj'} + \cdots + V_{12\cdots J}\right). \tag{3.20}$$

This shows that δ quantifies the fractional joint contribution of all biochemical factors to the response variance. Note that, if the response function is additive, then $\delta = 0$. If $\delta \simeq 0$, all joint contributions to the response variance will be negligible, whereas, appreciable values of δ indicate that these contributions may be significant. In the former case, Eq. (3.10) will be approximately satisfied and we can thus justify ranking the influential reactions by using the SESI's.

When $\delta \not\simeq 0$,[3] we need to investigate whether a biochemical factor contributes to the system response singularly, jointly, or both. To do so, we may calculate the difference

$$\eta_j := \tau_j - \sigma_j, \tag{3.21}$$

which, together with Eqs. (3.13) and (3.16), leads to

$$\eta_j = \frac{1}{V}\left(\sum_{j'\neq j}^{J} V_{jj'} + \cdots + V_{12\cdots J}\right). \tag{3.22}$$

[3]We use the notation $x \not\simeq 0$ to denote that x may take values that are sufficiently larger than zero.

Eqs. (3.1) and (3.22) imply that $0 \leq \eta_j \leq 1$, for every $j = 1, 2, \ldots, J$. According to Eq. (3.22), η_j quantifies the fractional contribution of the j^{th} biochemical factor to the response variance jointly with one or more other factors. For this reason, we refer to η_j as the *joint-effect* sensitivity index (JESI). If $\eta_j \simeq 0$, these contributions will be negligible, whereas, appreciable values of η_j indicate significant joint contributions. In the former case, we have $\tau_j \simeq \sigma_j$, by virtue of Eq. (3.21), which also implies that $\delta \simeq 0$, based on Eqs. (3.18) and (3.19). In this case, if $\sigma_j \simeq 0$, then $\tau_j \simeq 0$, and we may conclude that the j^{th} biochemical factor does not appreciably influence the system response (i.e., it is non-influential); whereas, if $\sigma_j \not\simeq 0$, then $\tau_j \not\simeq 0$, and we may conclude that the j^{th} biochemical factor influences the system response but mostly singularly. On the other hand, when $\eta_j \not\simeq 0$, the singular and joint contribution of the j^{th} biochemical factor to the response variance may be appreciable. In this case, if $\sigma_j \simeq 0$, we may conclude that the system response is not appreciably influenced by the j^{th} factor alone, but mostly by the j^{th} factor jointly with one or more other factors; whereas, if $\sigma_j \not\simeq 0$, we may conclude that the j^{th} biochemical factor influences the response function both singularly and jointly with one or more other reactions. We summarize these remarks in Table 3.1, which shows that the JESI's and SESI's are sufficient for proper classification of biochemical factors.

As we mentioned before, when the system response is not appreciably influenced by joint effects (i.e., when $\eta_j \simeq 0$, for $j = 1, 2, \ldots, J$), Eq. (3.10) is satisfied and we can use the SESI's (or the TESI's, since $\tau_j \simeq \sigma_j$ in this case) to rank the influential biochemical factors. When this is not true, we must identify an appropriate sensitivity index that we can use to rank the influential biochemical factors in the presence of joint

Table 3.1: Rules for interpreting the variance-based sensitivity indices.

	$\sigma_j \simeq 0$	$\sigma_j \not\simeq 0$
$\eta_j \simeq 0$	factor j does not appreciably influence the system response	factor j influences the system response mostly *singularly*
$\eta_j \not\simeq 0$	factor j influences the system response mostly *jointly*	factor j influences the system response both *singularly* and *jointly*

effects. To do so, we first need to determine the purpose of identifying a biochemical factor as "more influential" than another factor, and mathematically formalize what we mean by this comparison.

It is intuitive to believe that, to maintain robust behavior, a biochemical reaction system must control the standard chemical potentials associated with influential reactions and molecular species in a precise manner in order to reduce variations in its response. This observation motivates us to define a biochemical factor as being the *most* influential if, by fixing its value, the system response variance is, on the average, the *smallest* possible. Likewise, we may define the second most influential biochemical factor, and so on. Now, if we go back to Eq. (3.4), we have that $\mathrm{E}\big[\mathrm{Var}\big[R(\mathbf{U}) \mid U_j\big]\big]$ $= V - V_j$. This implies that, on the average, the smallest response variance is obtained by fixing the biochemical factor with the largest V_j value. As a consequence, we can still use the SESI's to rank the influential biochemical factors, even factors that

jointly contribute to the response variance. This ranking can be useful even in cases of appreciable joint effects, especially when the problem is to influence the response of a biochemical reaction system by only targeting biochemical factors one at a time. However, we must always keep in mind that a ranking based on SESI values considers only the singular contribution of each biochemical factor to the response variance and does not take into account joint effects.

When $\eta_j \not\simeq 0$, we may want to further investigate the joint influence of two biochemical factors j and j' on the system response. We can achieve this objective by employing the following index:

$$\upsilon_{jj'} := \frac{V_{jj'}}{V}, \quad j' \neq j, \tag{3.23}$$

which we refer to as the *pairwise-effect* sensitivity index (PESI). From Eq. (3.14) and Eqs. (3.21)–(3.23), we have that $0 \leq \upsilon_{jj'} \leq \eta_j \leq \tau_j \leq 1$, for $j' \neq j$. If $\upsilon_{jj'} \simeq 0$, we may conclude that the joint contribution of the biochemical factors j and j' to the system response is negligible, whereas, larger values of $\upsilon_{jj'}$ indicate stronger joint contributions.

In addition to ranking individual biochemical factors, it may also be desirable to rank pairs of factors based on their contribution to the response variance. From Eq. (3.5), we have that $\mathrm{E}\big[\mathrm{Var}\big[R(\mathbf{U}) \mid U_j, U_{j'}\big]\big] = V - (V_j + V_{j'} + V_{jj'})$. This implies that, on the average, by fixing two factors, the smallest response variance is obtained when fixing the factors with the largest $V_j + V_{j'} + V_{jj'}$ value. As a consequence, we can evaluate the influence of pairs of biochemical factors on the system response by

employing the following index:

$$\sigma_{jj'} := \frac{V_j + V_{j'} + V_{jj'}}{V} = \frac{\mathrm{Var}\big[\mathrm{E}\big[R(\mathbf{U}) \mid U_j, U_{j'}\big]\big]}{\mathrm{Var}\big[R(\mathbf{U})\big]}, \quad j' \neq j. \tag{3.24}$$

We refer to $\sigma_{jj'}$ as the *double-effect* sensitivity index (DESI). This index quantifies the average fractional reduction in the response variance when the two factors j and j' are fixed. From Eqs. (3.1) and (3.24), we have that $0 \leq \sigma_{jj'} \leq 1$, for $j' \neq j$. Thus, the DESI's are also normalized to take values between 0 and 1. Note also that

$$\sigma_{jj'} = \sigma_j + \sigma_{j'} + \upsilon_{jj'}, \quad j' \neq j. \tag{3.25}$$

Therefore, when $\upsilon_{jj'} = 0$, the DESI $\sigma_{jj'}$ is simply the sum of the SESI's of the two biochemical factors j and j'.

Finally, it is worthwhile noticing that, Eqs. (3.1), (3.16), and (3.23) imply that we can quantify the fractional contribution of all joint effects of order ≥ 3 to the response variance by means of

$$\gamma := 1 - \sum_{j=1}^{J} \sigma_j - \sum_{j=1}^{J} \sum_{j'>j}^{J} \upsilon_{jj'}. \tag{3.26}$$

When $\gamma \simeq 0$, these effects are negligible, in which case the use of indices σ_j, τ_j, η_j, and $\upsilon_{jj'}$ for sensitivity analysis will be sufficient. However, when $\gamma \not\simeq 0$, we may want to investigate higher-order joint effects of triplets, quadruples, etc. To do so, we would have to evaluate higher-order sensitivity indices, which would require additional computational resources (memory and CPU time). For this reason, we limit our analysis to first- and second-order effects. We expect that, in most cases of interest, these effects will provide a sufficient picture of the sensitivity properties of a biochemical reaction system.

We now summarize an algorithm for investigating the first- and second-order sensitivity properties of a biochemical reaction system by variance-based sensitivity analysis:

Initialization

1. Calculate the TESI's τ_j and the SESI's σ_j.

2. Calculate the JESI's $\eta_j = \tau_j - \sigma_j$.

3. Set a small threshold $\theta \ll 1$.

Classification

For $j = 1, 2, \ldots, J$:

4. If $\eta_j \leq \theta$ and $\sigma_j \leq \theta$, conclude that the j^{th} biochemical factor does not appreciably influence the system response.

5. If $\eta_j \leq \theta$ and $\sigma_j > \theta$, conclude that the j^{th} biochemical factor influences the system response mostly singularly.

6. If $\eta_j > \theta$ and $\sigma_j \leq \theta$, conclude that the j^{th} biochemical factor influences the system response mostly jointly with other factors.

7. If $\eta_j > \theta$ and $\sigma_j > \theta$, conclude that the j^{th} biochemical factor influences the system response both singularly and jointly with other factors.

Ranking

8. Use the SESI's to rank the influential factors, with the most influential factor being the one with the largest SESI value, the second most influential factor being the one with the second largest value, and so on.

9. If desired, calculate the PESI's and DESI's. Use the PESI's to investigate the contribution to the response variance of biochemical factors that influence the system response jointly with another factor. Use the DESI's to rank pairs of

biochemical factors, with the most influential pair being the one with the largest DESI value, the second most influential pair being the one with the second largest value, and so on.

3.2 Monte Carlo Estimation

The sensitivity indices derived in Section 3.1 cannot be computed analytically. However, a number of numerical techniques are available for their evaluation, with the most prominent ones based on Monte Carlo simulation [23,27]. In this subsection, we present a Monte Carlo method for estimating the variance-based sensitivity indices that uses a Latin hypercube sampling scheme [34, 66–68] to efficiently sample the random factors and reduce estimation variance. We will be referring to this technique as *Monte Carlo Latin Hypercube Sampling* (MC-LHS).

The MC-LHS method starts by forming two groups

$$\begin{array}{|cccc|}\hline u_1^{(1)} & u_2^{(1)} & \dots & u_J^{(1)} \\ u_1^{(2)} & u_2^{(2)} & \dots & u_J^{(2)} \\ \vdots & \vdots & & \vdots \\ u_1^{(L)} & u_2^{(L)} & \dots & u_J^{(L)} \\ \hline \end{array}$$

$$\begin{array}{|cccc|}\hline u_1^{(L+1)} & u_2^{(L+1)} & \dots & u_J^{(L+1)} \\ u_1^{(L+2)} & u_2^{(L+2)} & \dots & u_J^{(L+2)} \\ \vdots & \vdots & & \vdots \\ u_1^{(2L)} & u_2^{(2L)} & \dots & u_J^{(2L)} \\ \hline \end{array}$$

of $2L$ Latin hypercube samples of the statistically independent random factors $\mathbf{U} = \{U_1, U_2, \ldots, U_J\}$, where L is a given sample size.[4] The samples are drawn independently from the Gaussian probability densities of U_j, $j = 1, 2, \ldots, J$. In particular, when $U_j = Y_j^{\ddagger}$, the sample is drawn from a zero-mean Gaussian distribution with standard deviation $\lambda_j^{\ddagger}$, whereas, when $U_j = Y_j$, the sample is drawn from a zero-mean Gaussian distribution with standard deviation λ_j.

Subsequently, we group the samples together to form the following values for $\mathbf{U}$:

$$\mathbf{u}^{(l)} = \left\{u_1^{(l)}, u_2^{(l)}, \ldots, u_J^{(l)}\right\}, \quad l = 1, 2, \ldots, 2L$$

$$\mathbf{u}_j^{(l)} = \left\{u_1^{(L+l)}, \ldots, u_{j-1}^{(L+l)}, u_j^{(l)}, u_{j+1}^{(L+l)}, \ldots, u_J^{(L+l)}\right\}, \quad j = 1, 2, \ldots, J, \quad l = 1, 2, \ldots, L$$

$$\mathbf{u}_{(j)}^{(l)} = \left\{u_1^{(l)}, \ldots, u_{j-1}^{(l)}, u_j^{(L+l)}, u_{j+1}^{(l)}, \ldots, u_J^{(l)}\right\}, \quad j = 1, 2, \ldots, J, \quad l = 1, 2, \ldots, L.$$

We use these values, together with Eq. (2.21), to determine the reaction rate constants of the biochemical reaction system and evaluate the corresponding $2L(J+1)$ system responses $R(\mathbf{u}^{(l)})$, $R(\mathbf{u}_j^{(l)})$, and $R(\mathbf{u}_{(j)}^{(l)})$, by solving Eq. (2.2) and by using Eqs. (2.6) and (2.7).

We use the evaluated responses to compute the following Monte Carlo estimators of the response variances (see the Appendix at the end of this chapter for more details

[4]To draw a Latin hypercube sample $\{u_k^{(q)}, k = 1, 2, \ldots, K\}$, $q = 1, 2, \ldots, Q$, of K independent random variables U_k, $k = 1, 2, \ldots, K$, of size Q, we first draw KQ samples $\{\xi_k^{(q)}, k = 1, 2, \ldots, K, q = 1, 2, \ldots, Q\}$ independently from the uniform distribution on $[0, 1]$, and then set $u_k^{(q)} = F_k^{-1}\left[(\pi_k^{(q)} - 1 + \xi_k^{(q)})/Q\right]$, where $F_k[\cdot]$ is the cumulative distribution function of random variable U_k and $\{\pi_k^{(q)}, q = 1, 2, \ldots, Q\}$ is an independent random permutation of $\{1, 2, \ldots, Q\}$.

on the derivation):

$$\widehat{\mathrm{Var}_j}\big[R(\mathbf{U})\big] \;=\; \frac{1}{4L}\Big[\sum_{l=1}^{L} R^2(\mathbf{u}^{(l)}) + \sum_{l=1}^{L} R^2(\mathbf{u}^{(L+l)}) + \sum_{l=1}^{L} R^2(\mathbf{u}_j^{(l)}) + \sum_{l=1}^{L} R^2(\mathbf{u}_{(j)}^{(l)})\Big] - \widehat{\mathrm{E}}_j^2\big[R(\mathbf{U})\big], \tag{3.27}$$

$$\widehat{\mathrm{Var}_{jj'}}\big[R(\mathbf{U})\big] \;=\; \frac{1}{4L}\Big[\sum_{l=1}^{L} R^2(\mathbf{u}_j^{(l)}) + \sum_{l=1}^{L} R^2(\mathbf{u}_{(j)}^{(l)}) + \sum_{l=1}^{L} R^2(\mathbf{u}_{j'}^{(l)}) + \sum_{l=1}^{L} R^2(\mathbf{u}_{(j')}^{(l)})\Big] - \widehat{\mathrm{E}}_{jj'}^2\big[R(\mathbf{U})\big], \tag{3.28}$$

$$\widehat{\mathrm{Var}}\big[\mathrm{E}\big[R(\mathbf{U}) \mid U_j\big]\big] \;=\; \frac{1}{2L}\Big[\sum_{l=1}^{L} R(\mathbf{u}^{(l)})R(\mathbf{u}_j^{(l)}) + \sum_{l=1}^{L} R(\mathbf{u}^{(L+l)})R(\mathbf{u}_{(j)}^{(l)})\Big] - \widehat{\mathrm{E}}_j^2\big[R(\mathbf{U})\big], \tag{3.29}$$

$$\widehat{\mathrm{Var}}\big[\mathrm{E}\big[R(\mathbf{U}) \mid \mathbf{U}_{(j)}\big]\big] \;=\; \frac{1}{2L}\Big[\sum_{l=1}^{L} R(\mathbf{u}^{(l)})R(\mathbf{u}_{(j)}^{(l)}) + \sum_{l=1}^{L} R(\mathbf{u}^{(L+l)})R(\mathbf{u}_j^{(l)})\Big] - \widehat{\mathrm{E}}_j^2\big[R(\mathbf{U})\big], \tag{3.30}$$

$$\widehat{\mathrm{Var}}\big[\mathrm{E}\big[R(\mathbf{U}) \mid U_j, U_{j'}\big]\big] \;=\; \frac{1}{2L}\Big[\sum_{l=1}^{L} R(\mathbf{u}_j^{(l)})R(\mathbf{u}_{(j')}^{(l)}) + \sum_{l=1}^{L} R(\mathbf{u}_{j'}^{(l)})R(\mathbf{u}_{(j)}^{(l)})\Big] - \widehat{\mathrm{E}}_{jj'}^2\big[R(\mathbf{U})\big], \tag{3.31}$$

where

$$\widehat{\mathrm{E}}_j\big[R(\mathbf{U})\big] \;=\; \sqrt{\frac{1}{2L}\Big[\sum_{l=1}^{L} R(\mathbf{u}^{(l)})R(\mathbf{u}^{(L+l)}) + \sum_{l=1}^{L} R(\mathbf{u}_j^{(l)})R(\mathbf{u}_{(j)}^{(l)})\Big]}, \tag{3.32}$$

$$\widehat{\mathrm{E}}_{jj'}\big[R(\mathbf{U})\big] \;=\; \sqrt{\frac{1}{2L}\Big[\sum_{l=1}^{L} R(\mathbf{u}_j^{(l)})R(\mathbf{u}_{(j)}^{(l)}) + \sum_{l=1}^{L} R(\mathbf{u}_{j'}^{(l)})R(\mathbf{u}_{(j')}^{(l)})\Big]}. \tag{3.33}$$

Finally, we estimate the variance-based sensitivity indices by [recall Eqs. (3.15), (3.16),

(3.21), (3.24), and (3.25)]

$$\widehat{\tau}_j \;=\; 1 - \frac{\widehat{\mathrm{Var}}\big[\mathrm{E}\big[R(\mathbf{U}) \mid \mathbf{U}_{(j)}\big]\big]}{\widehat{\mathrm{Var}}_j\big[R(\mathbf{U})\big]}, \tag{3.34}$$

$$\widehat{\sigma}_j \;=\; \frac{\widehat{\mathrm{Var}}\big[\mathrm{E}\big[R(\mathbf{U}) \mid U_j\big]\big]}{\widehat{\mathrm{Var}}_j\big[R(\mathbf{U})\big]}, \tag{3.35}$$

$$\widehat{\eta}_j \;=\; \widehat{\tau}_j - \widehat{\sigma}_j, \tag{3.36}$$

$$\widehat{\sigma}_{jj'} \;=\; \frac{\widehat{\mathrm{Var}}\big[\mathrm{E}\big[R(\mathbf{U}) \mid U_j, U_{j'}\big]\big]}{\widehat{\mathrm{Var}}_{jj'}\big[R(\mathbf{U})\big]}, \quad j' \neq j, \tag{3.37}$$

$$\widehat{\upsilon}_{jj'} \;=\; \widehat{\sigma}_{jj'} - \widehat{\sigma}_j - \widehat{\sigma}_{j'}, \quad j' \neq j. \tag{3.38}$$

The above estimators are modified versions of the ones proposed previously by Saltelli [31]. These modifications are important in order to guarantee that the estimated variances satisfy the following necessary conditions for any number of Monte Carlo samples (whose proofs are provided in the Appendix at the end of this chapter):

C.1: $\widehat{\mathrm{Var}}_j\left[R(\mathbf{U})\right] \geq 0$ and $\widehat{\mathrm{Var}}_{jj'}\left[R(\mathbf{U})\right] \geq 0$.

C.2: If U_j is fixed, then $\widehat{\mathrm{Var}}\left[\mathrm{E}\left[R(\mathbf{U}) \mid U_j\right]\right] = 0$, $\widehat{\mathrm{Var}}\left[\mathrm{E}\left[R(\mathbf{U}) \mid \mathbf{U}_{(j)}\right]\right] = \widehat{\mathrm{Var}}_j\left[R(\mathbf{U})\right]$.

C.3: If $\mathbf{U}_{(j)}$ is fixed, then $\widehat{\mathrm{Var}}\left[\mathrm{E}\left[R(\mathbf{U}) \mid U_j\right]\right] = \widehat{\mathrm{Var}}_j\left[R(\mathbf{U})\right]$, $\widehat{\mathrm{Var}}\left[\mathrm{E}\left[R(\mathbf{U}) \mid \mathbf{U}_{(j)}\right]\right] = 0$.

C.4: $\widehat{\mathrm{Var}}\left[\mathrm{E}\left[R(\mathbf{U}) \mid U_j\right]\right] + \widehat{\mathrm{Var}}\left[\mathrm{E}\left[R(\mathbf{U}) \mid \mathbf{U}_{(j)}\right]\right] \leq \widehat{\mathrm{Var}}_j\left[R(\mathbf{U})\right]$.

C.5: If U_j and $U_{j'}$ are fixed, then $\widehat{\mathrm{Var}}\left[\mathrm{E}\left[R(\mathbf{U}) \mid U_j, U_{j'}\right]\right] = 0$.

C.6: If $\mathbf{U}_{(j,j')}$ is fixed, then $\widehat{\mathrm{Var}}\left[\mathrm{E}\left[R(\mathbf{U}) \mid U_j, U_{j'}\right]\right] = \widehat{\mathrm{Var}}_{jj'}\left[R(\mathbf{U})\right]$.

C.7: $\widehat{\mathrm{Var}}\left[\mathrm{E}\left[R(\mathbf{U}) \mid U_j, U_{j'}\right]\right] + \widehat{\mathrm{Var}}\left[\mathrm{E}\left[R(\mathbf{U}) \mid \mathbf{U}_{(j,j')}\right]\right] \leq \widehat{\mathrm{Var}}_{jj'}\left[R(\mathbf{U})\right]$.

Note that, in Conditions C.6 and C.7, $\mathbf{U}_{(j,j')}$ is the set of all factors excluding U_j and $U_{j'}$.

Condition C.1 guarantees that the two estimators for the response variance are nonnegative. Estimators for the conditional response variances $\mathrm{Var}\big[\mathrm{E}\big[R(\mathbf{U}) \mid U_j\big]\big]$,

$\mathrm{Var}\big[\mathrm{E}\big[R(\mathbf{U}) \mid \mathbf{U}_{(j)}\big]\big]$, and $\mathrm{Var}\big[\mathrm{E}\big[R(\mathbf{U}) \mid U_j, U_{j'}\big]\big]$ must also be nonnegative. However, this is not necessarily true for the previous estimators. We could derive nonnegative variance estimators by employing the standard Monte Carlo formulas

$$\widehat{\mathrm{Var}}\,[U] = \frac{1}{L}\sum_{l=1}^{L}(u^{(l)} - \widehat{\mathrm{E}}\,[U])^2, \qquad \text{where} \quad \widehat{\mathrm{E}}\,[U] = \frac{1}{L}\sum_{l=1}^{L} u^{(l)}, \tag{3.39}$$

but using these formulas results in an unnecessarily large number of system response evaluations [31]. Note however that, if the variance estimators are nonnegative, then condition C.4 guarantees that $0 \leq \widehat{\sigma}_j \leq 1$, $0 \leq \widehat{\tau}_j \leq 1$, and $\widehat{\sigma}_j \leq \widehat{\tau}_j$, as expected. This implies that $0 \leq \widehat{\eta}_j := \widehat{\tau}_j - \widehat{\sigma}_j \leq 1$. Moreover, condition C.7 guarantees that $0 \leq \widehat{\sigma}_{jj'} \leq 1$, which also implies that $\widehat{\upsilon}_{jj'} \leq 1$. On the other hand, conditions C.2 and C.3 guarantee that, when U_j is fixed, then $\widehat{\sigma}_j = \widehat{\tau}_j = 0$, whereas, when $\mathbf{U}_{(j)}$ is fixed, then $\widehat{\sigma}_j = \widehat{\tau}_j = 1$. Finally, conditions C.5 and C.6 guarantee that, when U_j and $U_{j'}$ are fixed, then $\widehat{\sigma}_{jj'} = 0$, whereas, when $\mathbf{U}_{(j,j')}$ is fixed, then $\widehat{\sigma}_{jj'} = 1$. In our experience, by using the previous estimators, instead of the ones suggested in [31], we obtain a more efficient numerical implementation of the sensitivity analysis methodology discussed in Section 3.1, which, with fewer Monte Carlo samples, produces estimates of the underlying sensitivity indices consistent with all necessary conditions and constraints.

3.3 Numerical Results

We now use the MC-LHS technique discussed in the previous subsection to estimate the variance-based sensitivity indices, given by Eqs. (3.34)–(3.38), for the logarithms of the duration, integrated response and strength of the ERK-PP concentration profile in the MAPK biochemical reaction model depicted in Fig. 2.1. We consider three strategies

for sensitivity analysis: (a) reaction-oriented sensitivity analysis (ROSA), (b) species-oriented sensitivity analysis (SOSA), and (c) reaction-oriented/species-oriented sensitivity analysis (ROSOSA). ROSA investigates only the effects of fluctuations in the standard chemical potentials of the activated complexes, whereas, SOSA investigates only the effects of fluctuations in the standard chemical potentials associated with the molecular species. On the other hand ROSOSA investigates the sensitivity behavior of a biochemical reaction system under both types of fluctuations.

ROSA and SOSA can be useful in drug-design problems in which pharmacological control of the response characteristics of a biochemical reaction system is of interest. ROSA can be used when the objective is to modify the response of a biochemical reaction system by pharmacologically targeting selected enzymes responsible for catalyzing influential reactions, whereas, SOSA can be used when the objective is to modify the system response by altering the kinetic properties of selected influential molecular species. On the other hand, ROSOSA provides a more general approach to sensitivity analysis. As a matter of fact, we can use ROSOSA to obtain a complete picture of the sensitivity properties of a biochemical reaction system both with respect to the underlying reactions and reactant molecular species. In this subsection, we illustrate the use of ROSOSA for identifying the reaction rate constants responsible for influencing the system response. This task is important in reverse engineering problems, where the objective is to estimate the reaction rate constants of a biochemical reaction system from available data. Use of ROSOSA may help us to focus our estimation effort on "influential" reaction rate constants, whose values must be determined with high accuracy, and ignore the remaining "non-influential" reaction rate constants, whose

exact values are of no particular interest.

Although, in general, the standard deviations associated with the standard chemical potentials of the activated complexes and molecular species depend on m and n, respectively, this dependence may not be useful in practice, since it is difficult to obtain information about the fluctuation level of each individual standard chemical potential. For this reason, we assume here that $\lambda_m^\ddagger = \lambda^\ddagger$, $\lambda_n = \lambda$, and consider $\lambda^\ddagger$ and λ as two "user-defined" parameters that control the "scale" of sensitivity analysis. Small values of $\lambda^\ddagger$ and λ correspond to "local" sensitivity analysis, associated with small fluctuations in the standard chemical potentials about their nominal values, whereas, large values of $\lambda^\ddagger$ and λ correspond to "global" sensitivity analysis, associated with large fluctuations in the standard chemical potentials. Note that, even under this simplification, when using SOSA (in which case $\lambda^\ddagger = 0, \lambda \neq 0$) or ROSOSA (in which case $\lambda^\ddagger, \lambda \neq 0$), the size of variations applied on the reaction rates will not be uniform, due to Eq. (2.23).

In this subsection, we investigate the sensitivity properties of MAPK by implementing the classification/ranking steps presented in Section 3.1 with threshold $\theta = 0.1$, for the ROSA and SOSA, and $\theta = 0.05$, for the ROSOSA, which corresponds to 10% and 5% of the maximum attainable JESI and SESI values, respectively. We take the ROSOSA threshold value to be half of the one used in ROSA and SOSA, since, in the former case, the response variance is distributed among two types of biochemical factors (i.e., among the standard chemical potentials of the activated complexes and the standard chemical potentials of the molecular species), as opposed to the latter cases in which the response variance is distributed among only one type of biochemical factors. Choosing a threshold value is a relatively easy task in variance-based sensitivity analy-

sis techniques, since the indices are normalized to take values between 0 and 1. In the following, we estimate the variance-based sensitivity indices associated with the duration, integrated response, and strength of the ERK-PP concentration profile, as defined by Eqs. (2.6) and (2.7). We do this by considering the dynamic behavior of MAPK within a time frame of 6 hours ($t_{\max} = 360$ min), by setting $\varepsilon = 0.05$ in Eq. (2.7), and by employing the MC-LHS estimators presented in Section 3.2 with $L = 6{,}000$.

In Fig. 3.1, we depict the ROSA results for MAPK at three different fluctuation levels of the standard chemical potentials of the activated complexes.[5] In the case of duration, the estimated values of γ turn out to be all zero, which implies that there is no appreciable fractional contribution to the response variances from high-order (≥ 3) joint effects. The same is true in the case of integrated response with $\lambda^{\ddagger} = 0.1, 0.2$ and strength with $\lambda^{\ddagger} = 0.1$. However, when $\lambda^{\ddagger} = 0.2$, the estimated γ value for strength is 0.06; when $\lambda^{\ddagger} = 0.4$, the estimated γ values for integrated response and strength are 0.06 and 0.14, respectively, which indicate emergence of high-order (≥ 3) joint effects for larger values of $\lambda^{\ddagger}$, although these effects are still relatively small. A closer look at the results depicted in Fig. 3.1 indicates that the integrated response and strength may be subject to second-order joint effects, since, when $\lambda^{\ddagger} = 0.4$, some JESI values associated with these response characteristics are above the threshold. As a matter of fact the estimated value of δ for duration is relatively small ($\delta \leq 0.15$), but the value associated with the integrated response when $\lambda^{\ddagger} = 0.4$ and the value associated with the strength when $\lambda^{\ddagger} = 0.2, 0.4$ are both large. These values have been estimated to

[5] Eq. (2.22) implies that an increase in the chemical potential value of the activated complex of a reaction one standard deviation from the (zero) mean produces a variation in the reaction rate constant that amounts to $100e^{\lambda^{\ddagger}}\%$ of the nominal value. As a consequence, the values of $\lambda^{\ddagger}$ considered in this chapter correspond to perturbing the reaction rate constants about 10%, 20%, and 50% of their nominal values, when $\lambda^{\ddagger} = 0.1$, $\lambda^{\ddagger} = 0.2$, and $\lambda^{\ddagger} = 0.4$, respectively.

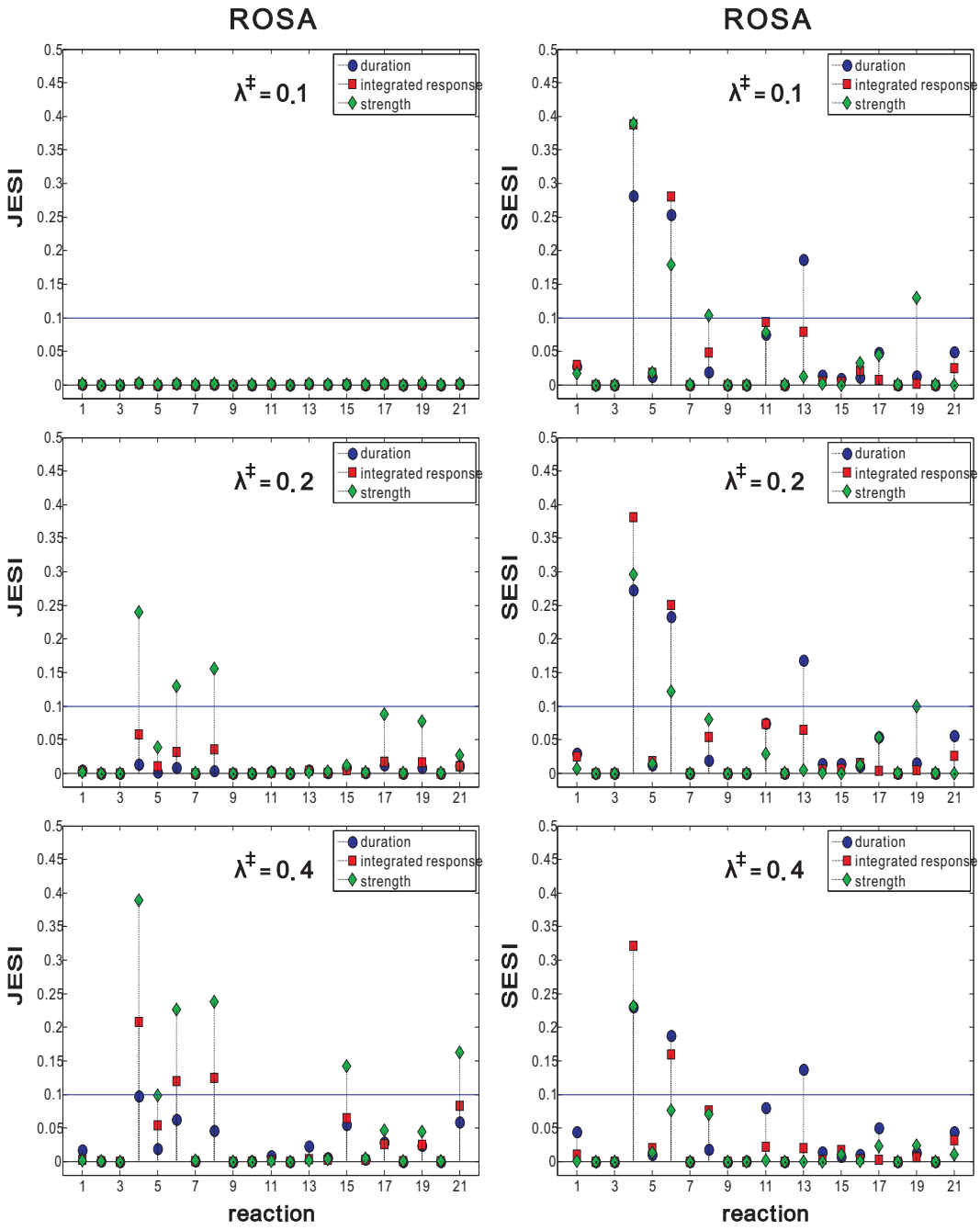

Figure 3.1: ROSA results for the MAPK signaling cascade at three different fluctuation levels with $\lambda^{\ddagger} = 0.1, 0.2, 0.4$.

be 0.30, 0.28 and 0.53, respectively.

In all three cases depicted in Fig. 3.1, the duration is *singularly* influenced by the same three reactions 4 (Raf* + Pho1 $\rightleftharpoons$ Raf*-Pho1), 6 (MEK + Raf* $\rightleftharpoons$ MEK-Raf*), and 13 (MEK-P-Pho2 $\rightarrow$ MEK + Pho2), with reaction 4 being the most influential

and reaction 13 being the least influential. As a matter of fact, the estimated SESI values indicate that these three reactions account for about 72%, 67%, and 56% of the duration variance when $\lambda^{\ddagger} = 0.1, 0.2$, and 0.4, respectively. However, the results depicted in Fig. 3.1 indicate more complex sensitivity behaviors for the integrated response and strength.

When $\lambda^{\ddagger} = 0.1, 0.2$, the ROSA results indicate that the integrated response is influenced *singularly* by reaction 4 (Raf* + Pho1 $\rightleftharpoons$ Raf*-Pho1) and reaction 6 (MEK + Raf* $\rightleftharpoons$ MEK-Raf*), with reaction 4 being the most influential and reaction 6 being the second most influential. As a matter of fact, the estimated SESI values indicate that these two reactions account for about 67% and 63% of the integrated response variance when $\lambda^{\ddagger} = 0.1$ and 0.2, respectively.

When $\lambda^{\ddagger} = 0.4$, the integrated response is influenced both *singularly* and *jointly* by reaction 4 (Raf* + Pho1 $\rightleftharpoons$ Raf*-Pho1) and reaction 6 (MEK + Raf* $\rightleftharpoons$ MEK-Raf*), with reaction 8 (MEK-P + Raf* $\rightleftharpoons$ MEK-P-Raf*) influencing the integrated response only *jointly*. Inspection of the estimated SESI values and the estimated PESI values (data not shown) reveals that the singular and pairwise effects associated with these three reactions account for about 77% of the integrated response variance, with 69% of this amount being attributed to singular and pairwise effects solely among reactions 4 and 6.

When $\lambda^{\ddagger} = 0.1$, the ROSA results indicate that the strength is influenced *singularly* by reaction 4 (Raf* + Pho1 $\rightleftharpoons$ Raf*-Pho1), reaction 6 (MEK + Raf* $\rightleftharpoons$ MEK-Raf*), reaction 19 (ERK-PP-Pho3 $\rightarrow$ ERK-P + Pho3), and reaction 8 (MEK-P + Raf* $\rightleftharpoons$ MEK-P-Raf*), with reaction 4 being the most influential and reaction 8 being the

least influential. As a matter of fact, the estimated SESI values indicate that these four reactions account for about 80% of the strength variance.

When $\lambda^{\ddagger} = 0.2$, the strength is being influenced both *singularly* and *jointly* by reaction 4 (Raf* + Pho1 $\rightleftharpoons$ Raf*-Pho1) and reaction 6 (MEK + Raf* $\rightleftharpoons$ MEK-Raf*), with reaction 8 (MEK-P + Raf* $\rightleftharpoons$ MEK-P-Raf*) influencing the strength only *jointly*. Inspection of the estimated PESI values (data not shown) reveals that the pairwise influence of reaction 4 on the strength is mostly with reactions 6, 19, 17, and 8. It turns out that the singular and pairwise effects associated with reactions 4, 6, and 8 account for about 69% of the strength variance, with 86% of this amount being attributed to singular and pairwise effects solely among reactions 4, 6, and 8. On the other hand, inspection of the estimated DESI values (data not shown) reveals that the pairs $(4-6)$, $(4-8)$, and $(6-8)$ account for about 48%, 40%, and 21% of the strength variance, respectively.

Finally, when $\lambda^{\ddagger} = 0.4$, the strength is influenced both *singularly* and *jointly* only by reaction 4 (Raf* + Pho1 $\rightleftharpoons$ Raf*-Pho1), with reaction 6 (MEK + Raf* $\rightleftharpoons$ MEK-Raf*), reaction 8 (MEK-P + Raf* $\rightleftharpoons$ MEK-P-Raf*), reaction 15 (ERK-MEK-PP $\rightarrow$ ERK-P + MEK-PP), and reaction 21 (ERK-P-Pho3 $\rightarrow$ ERK + Pho3) influencing the strength only *jointly*. Inspection of the estimated SESI values and the estimated PESI values (data not shown) reveals that the singular and pairwise effects associated with reactions 4, 6, and 8 account for about 74% of the strength variance, with 74% of this amount being attributed to singular and pairwise effects solely among reactions 4, 6, and 8. On the other hand, inspection of the estimated DESI values (data not shown) reveals that the pairs $(4-6)$, $(4-8)$, and $(6-8)$ account for about 38%, 37%, and

Table 3.2: MAPK ROSA results obtained by MC-LHS.

No.	Reaction	D	I	S
4	Raf* + Pho1 $\rightleftharpoons$ Raf*-Pho1	•	•	•
6	MEK + Raf* $\rightleftharpoons$ MEK-Raf*	•	•	•
8	MEK-P + Raf* $\rightleftharpoons$ MEK-P-Raf*			•
13	MEK-P-Pho2 $\rightarrow$ MEK + Pho2	•		

17% of the strength variance, respectively.

As a consequence of the previous results, we may conclude that the duration is influenced by reactions 4, 6, and 13; the integrated response is predominantly influenced by reactions 4 and 6; whereas, the strength is predominantly influenced by reactions 4, 6, and 8; see Table 3.2.

In Fig. 3.2, we depict the SESI values evaluated by means of Eq. (3.17), with the response derivatives being approximated by symmetric finite-differences around the nominal reaction rate values, as well as the SESI values estimated by MC-LHS. Note that Eq. (3.17) provides a reasonable approximation of the estimated MC-LHS SESI values when $\lambda^{\ddagger} = 0.1$. The derivative-based SESI values correctly predict that, when $\lambda^{\ddagger} = 0.1$, reactions 4, 6, and 13 are influential for the duration, reactions 4 and 6 are influential for the integrated response, whereas, reactions 4, 6, 19, and 8 are influential for the strength. They predict that reactions 4, 6, and 13 account for 73% of the duration variance, as compared to 72% predicted by MC-LHS. They also predict that reactions 4 and 6 account for 66% of the integrated response variance, as compared

to 67% predicted by MC-LHS. Moreover, these values predict that reactions 4, 6, 19, and 8 account for 79% of the strength variance, as compared to 80% predicted by MC-LHS. However, it is clear from the results depicted in Fig. 3.2 that the accuracy of the derivative-based SESI values decreases as $\lambda^{\ddagger}$ increases. As a matter of fact, the prediction that reactions 4, 6, and 13 account for 73% of the duration variance is inaccurate when compared to 67% and 56% predicted by MC-LHS for $\lambda^{\ddagger} = 0.2$ and $\lambda^{\ddagger} = 0.4$, respectively. The same is true for the integrated response and strength variances. Moreover, by using the derivative-based SESI values, we cannot detect the diminishing role of reaction 19 and the emergence of reaction 8 as a key influential factor for the integrated response. These deficiencies are expected, since the derivative-based SESI approximation, given by Eq. (3.17), cannot account for joint effects, which become prominent at increasing levels of standard chemical potential fluctuations. Therefore, it becomes clear that special care should be exercised when employing the approximation given by Eq. (3.17) for sensitivity analysis, whose use must be limited to problems with negligible joint effects.[6]

In Fig. 3.3, we depict the SOSA results at three different fluctuation levels of the standard chemical potentials of the molecular species. When $\lambda = 0.1$, the estimated values of γ turn out to be zero for all the three response characteristics, which implies that there is no appreciable fractional contribution to the response variances from high-order (≥ 3) joint effects. The same is true in the case of duration and integrated response with $\lambda^{\ddagger} = 0.2$. However, when $\lambda = 0.2$, the estimated γ value for strength

[6] In this case, the TESI is approximately equal to the SESI and, therefore, the JESI is approximately zero. Note also that Eq. (3.17) implies that the SESI is independent of the standard deviations λ_j, when their values are all equal.

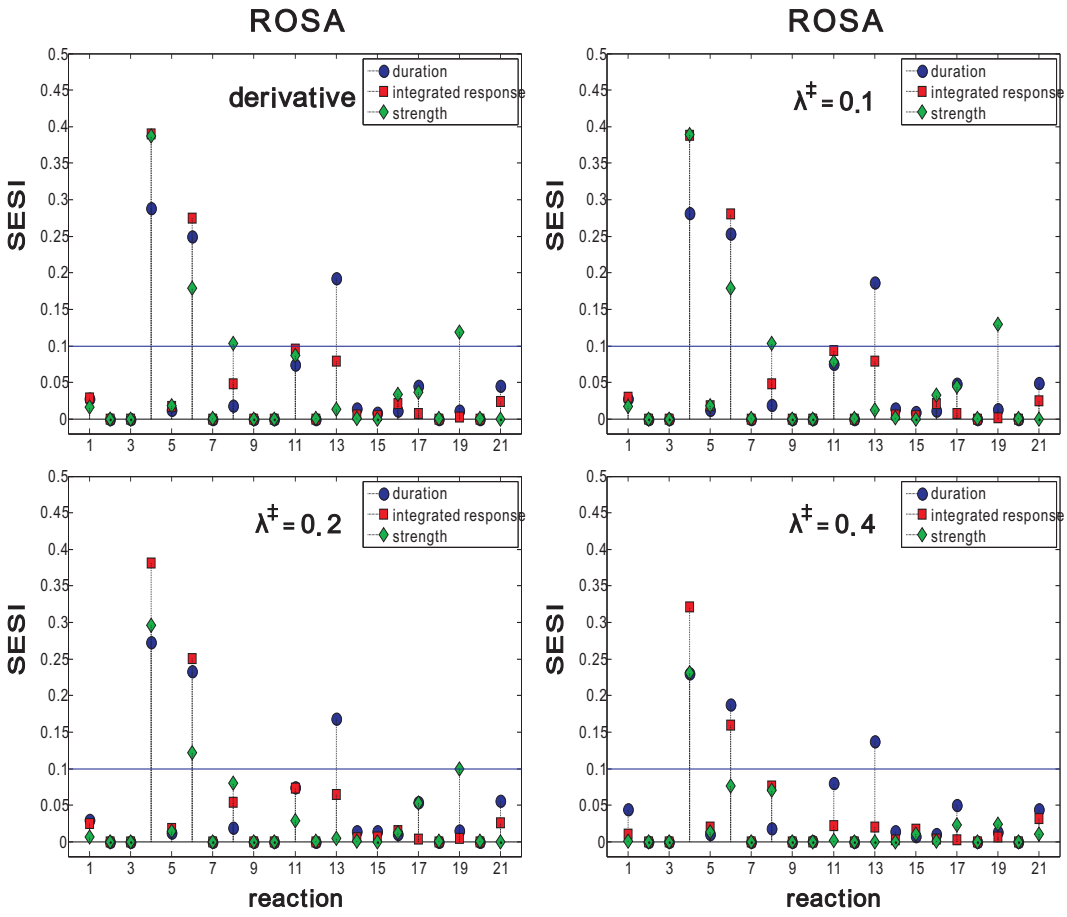

Figure 3.2: Derivative-based vs. MC-LHS estimation of SESI values associated with ROSA, when $\lambda^{\ddagger} = 0.1, 0.2, 0.4$.

is 0.06; when $\lambda = 0.4$, the estimated γ values are 0.07, 0.05, and 0.13, for duration, integrated response, and strength, respectively, which indicate emergence of high-order (≥ 3) joint effects for larger values of λ, although these effects are still relatively small. A closer look at the results indicates that all three response characteristics may be subject to second-order joint effects, since, when $\lambda = 0.4$, a JESI value associated with the duration is above the 10% threshold, whereas, several JESI values associated with the integrated response and strength are above the 10% threshold. As a matter of fact the estimated values of δ for all response characteristics are small when $\lambda^{\ddagger} = 0.1$ ($\delta \leq 0.02$), but when the factor fluctuation level increases, larger estimated δ values

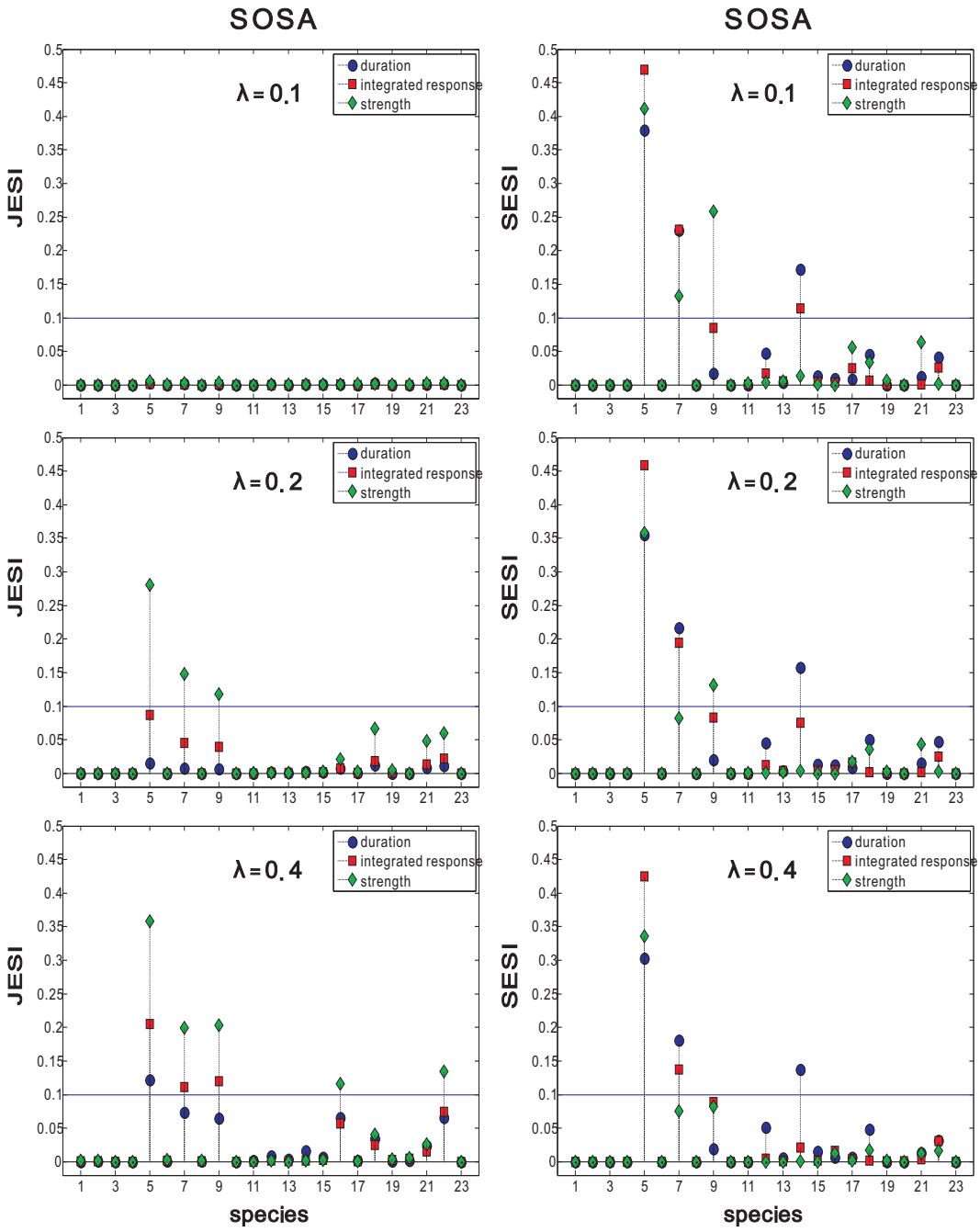

Figure 3.3: SOSA results for the MAPK signaling cascade at three different fluctuation levels with $\lambda = 0.1, 0.2, 0.4$.

indicate emergence of substantial joint effects, especially for strength, whose δ values have been estimated to be 0.32 and 0.44 when $\lambda = 0.2$ and $\lambda = 0.4$, respectively.

In all three cases depicted in Fig. 3.3, the duration is *singularly* influenced by the same three molecular species, 5 (Pho1), 7 (MEK), and 14 (MEK-P-Pho2), with

species 5 being the most influential and species 14 being the least influential. As a matter of fact, the estimated SESI values indicate that these three species account for about 78%, 73%, and 62% of the duration variance when $\lambda = 0.1, 0.2$, and 0.4, respectively. When $\lambda = 0.4$, the duration is influenced by species 5 both *singularly* and *jointly*. Inspection of the estimated PESI values (data not shown) reveals that the pairwise influence of species 5 on the duration is mostly with species 7, 14, and 18 (ERK-P-MEK-PP). However, these pairwise influences contribute only 3.4% to the duration variance. On the other hand, inspection of the estimated DESI values (data not shown) reveals that the pairs $(5-7)$, $(5-14)$, and $(7-14)$ account for about 50%, 45%, and 33% of the duration variance, respectively.

When $\lambda^{\ddagger} = 0.1$, the SOSA results indicate that the integrated response is influenced *singularly* by species 5 (Pho1), 7 (MEK), and 14 (MEK-P-Pho2), with species 5 being the most influential and species 14 being the least influential. As a matter of fact, the estimated SESI values indicate that these three species account for about 81% of the integrated response variance when $\lambda^{\ddagger} = 0.1$.

When $\lambda^{\ddagger} = 0.2$, species 14 does not influence the integrated response anymore. In this case, the integrated response is influenced *singularly* only by species 5 and 7. Inspection of the estimated SESI values reveals that their singular effects account for about 65% of the integrated response variance.

When $\lambda^{\ddagger} = 0.4$, the integrated response is influenced both *singularly* and *jointly* by species 5 and 7, with species 9 (MEK-P) influencing the integrated response only *jointly*. Inspection of the estimated PESI values (data not shown) reveals that the pairwise influence of species 5 on the integrated response is mostly with species 7, 9, 18

(ERK-P-MEK-PP), and 22 (ERK-P-Pho3), whereas, species 7 influences the strength jointly with species 22. It turns out that the singular and pairwise effects associated with species 5, 7, and 9 account for about 84% of the integrated response variance, with 73% of this amount being attributed to singular and pairwise effects solely among species 5 and 7.

When $\lambda = 0.1$, the strength is being influenced *singularly* by species 5 (Pho1), 9 (MEK-P), and 7 (MEK), with species 5 being the most influential and species 7 being the least influential. The estimated SESI values indicate that these three species account for 80% of the strength variance.

When $\lambda = 0.2$, the strength is being influenced both *singularly* and *jointly* by species 5 and 9, whereas, species 7 influences the strength only *jointly*. Inspection of the estimated PESI values (data not shown) reveals that the pairwise influence of species 5 on the strength is mostly with species 7, 9, 16 (ERK-MEK-PP), 18 (ERK-P-MEK-PP), and 22 (ERK-P-Pho3). It turns out that the singular and pairwise effects associated with species 5 7, and 9 account for about 81% of the integrated response variance, with 84% of this amount being attributed to singular and pairwise effects solely among species 5, 7, and 9. On the other hand, inspection of the estimated DESI values (data not shown) reveals that the pairs $(5-9)$, $(5-7)$, and $(7-9)$ account for about 52%, 50%, and 23% of the strength variance, respectively.

Finally, when $\lambda = 0.4$, the strength is being influenced both *singularly* and *jointly* only by species 5, with species 9, 7, 16 (ERK-MEK-PP), and 22 (ERK-P-Pho3), influencing the strength only *jointly*. Inspection of the estimated PESI values (data not shown) reveals that the pairwise influence of species 5 on the strength is mostly

Table 3.3: MAPK SOSA results by MC-LHS.

No.	Molecular Species	D	I	S
5	Pho1	•	•	•
7	MEK	•	•	•
9	MEK-P			•
14	MEK-P-Pho2	•		

with species 7, 9, 16 (ERK-MEK-PP), 18 (ERK-P-MEK-PP), and 22 (ERK-P-Pho3), whereas, species 9 influences the strength jointly with species 16, and species 7 influences the strength jointly with species 22. It turns out that the singular and pairwise effects associated with species 5, 7, and 9 account for about 79% of the strength variance, with the singular and pairwise effects solely among species 5, 7, and 9 accounting for about 82% of that amount. Inspection of the estimated DESI values (data not shown) reveals that the pairs $(5-7)$, $(5-9)$, and $(7-9)$, account for about 49%, 49%, and 16% of the strength variance, respectively.

As a consequence of the previous results, we may conclude that the duration is influenced by species 5, 7, and 14; the integrated response is predominantly influenced by species 5 and 7; whereas, the strength is predominantly influenced by species 5, 7, and 9.

The results obtained by ROSOSA, with $\lambda^{\ddagger} = \lambda = 0.1, 0.2, 0.4$, perfectly agree with the previous conclusions and reveal no additional sensitivity behavior (data not shown). However, we can also use ROSOSA to identify the most influential reaction rates in

a biochemical reaction system. Indeed, from Eq. (2.21), note that the reaction rate constants K_{2m-1} and K_{2m} of the m^{th} forward and reverse reactions are influenced by the biochemical factors U_m and U_{M+n}, $n = 1, 2, \ldots, N$, through the zero-mean Gaussian random variables G_m and G'_m, respectively, given by Eq. (2.22). As a consequence, we can quantify the influence of the rate constants K_{2m-1} and K_{2m} on the system response by means of the *net* fractional contribution of the biochemical factors U_m and U_{M+n}, $n = 1, 2, \ldots, N$, on the system response. This leads to the following sensitivity indices:

$$\begin{aligned} g_{2m-1} &= \sigma_m + \sum_{n=1}^{N} \text{sgn}(\nu_{nm})\sigma_n + \sum_{n=1}^{N} \text{sgn}(\nu_{nm})\upsilon_{mn} + \sum_{n=1}^{N}\sum_{n'=n+1}^{N} \text{sgn}(\nu_{nm})\text{sgn}(\nu_{n'm})\upsilon_{nn'} \\ g_{2m} &= \sigma_m + \sum_{n=1}^{N} \text{sgn}(\nu'_{nm})\sigma_n + \sum_{n=1}^{N} \text{sgn}(\nu'_{nm})\upsilon_{mn} + \sum_{n=1}^{N}\sum_{n'=n+1}^{N} \text{sgn}(\nu'_{nm})\text{sgn}(\nu'_{n'm})\upsilon_{nn'}, \end{aligned} \tag{3.40}$$

where σ_m is the SESI of the m^{th} reaction, σ_n is the SESI of the n^{th} species, υ_{mn} is the PESI between reaction m and species n, $\upsilon_{nn'}$ is the PESI between species n and n', and $\text{sgn}(\cdot)$ is the sign function [i.e., $\text{sgn}(a) = 1$, if $a > 0$, whereas, $\text{sgn}(a) = 0$, if $a = 0$]. We refer to g_{2m-1} and g_{2m} as the *group-effect* sensitivity indices (GESI) of the forward and reverse m^{th} reaction, respectively. By using these indices, we can say that a reaction rate is *most* influential if the corresponding GESI value is the largest one. Likewise, we may define the second most influential rate constant, and so on.

In Fig. 3.4, we depict the ROSOSA results for the reaction rate constants at three different fluctuation levels of the standard chemical potentials. It is clear from these results that only a small fraction of the reaction rate constants appreciably influence the duration, integrated response, and strength of ERK-PP. We summarize these reactions in Table 3.4, which depicts only the reaction rates that, for a given response

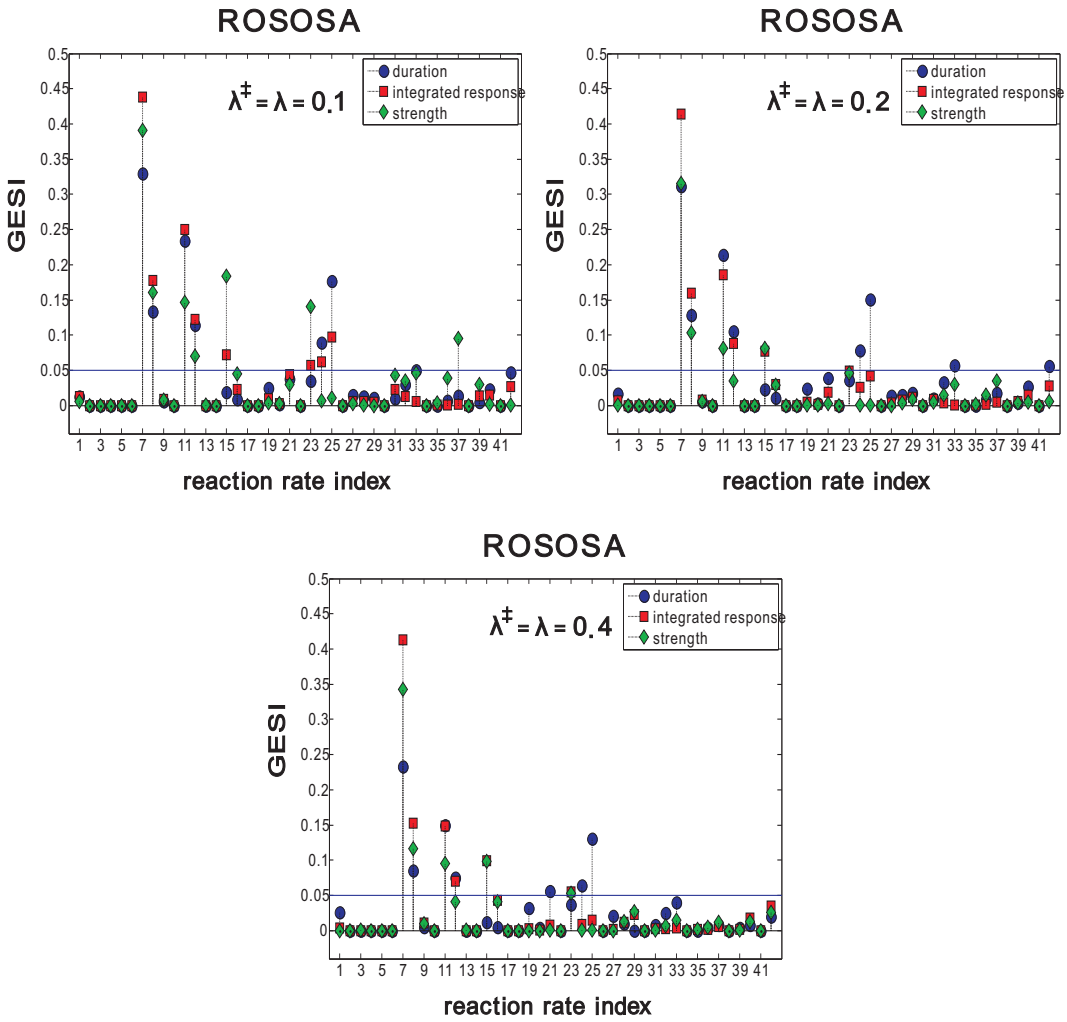

Figure 3.4: ROSOSA reaction rate results for the MAPK signaling cascade and for three different fluctuation levels with $\lambda^{\ddagger} = \lambda = 0.1, 0.2, 0.4$. The GESI values associated with the reaction rate constants k_2, k_{10}, k_{14}, k_{18}, k_{22}, k_{26}, k_{30}, k_{34}, k_{38}, and k_{41} are not calculated, since these constants are zero (they correspond to the reverse portion of the irreversible reactions $1, 5, 7, 9, 11, 13, 15, 17, 19$, and 21).

characteristic, are consistently classified as being influential at all three fluctuation levels. The results indicate that the duration is influenced by the reaction rate constants k_7, k_8, k_{11}, k_{12}, k_{24}, and k_{25}. Four of these rate constants, namely k_7, k_8, k_{11}, and k_{12}, influence the integrated response as well, and three of these rate constant, namely k_7, k_8, and k_{11}, influence the strength as well. The integrated response is also influenced by k_{15}, whereas, the strength is also influenced by k_{15} and k_{23}. From Fig. 3.4, the most

Table 3.4: MAPK ROSOSA results.

Rate	Reaction	D	I	S
k_7	Raf* + Pho1 → Raf*-Pho1	•	•	•
k_8	Raf*-Pho1 → Raf* + Pho1	•	•	•
k_{11}	MEK+Raf* → MEK-Raf*	•	•	•
k_{12}	MEK-Raf* → MEK+Raf*	•	•	
k_{15}	MEK-P+Raf* → MEK-P-Raf*		•	•
k_{23}	MEK-P+Pho2 → MEK-P-Pho2			•
k_{24}	MEK-P-Pho2 → MEK-P+Pho2	•		
k_{25}	MEK-P-Pho2 → MEK+Pho2	•		

influential rate constant for all the three characteristics is k_7.

3.4 Discussion

Our ROSA sensitivity analysis results summarized in Table 3.2 indicate that the binding and unbinding of the active version Raf* of the Raf kinase with its inactivator phosphatase Pho1 is the reaction that most influences the duration, integrated response and strength of ERK-PP activity in the MAPK signaling cascade. The second most influential reaction turns out to be the binding and unbinding of Raf* with MEK. The duration is also influenced by the dephosphorylation of the phosphorylated version MEK-P of MEK by the phosphatase Pho2, whereas, the strength is also influenced by

the binding and unbinding of Raf* with MEK-P. On the other hand, the SOSA sensitivity results summarized in Table 3.3 indicate that the phosphatase Pho1, associated with Raf inactivation, and the kinase MEK are two very important molecular species that are responsible for influencing the duration, integrated response and strength of ERK-PP activity in MAPK signaling. Interestingly, the Food and Drug Administration (FDA) has recently approved the use of Sorafenib for the treatment of kidney and liver cancer. This drug is a small molecular inhibitor that targets Raf (as well as other kinases) and induces anti-proliferative and proapoptotic effects by influencing ERK activity [69]. Another potent inhibitor of oncogenic B-Raf kinase activity, PLX4032, is currently under clinical trials for the treatment of late-stage melanoma. It has been demonstrated to selectively block the MAPK pathway in B-Raf mutant cells and to cause programmed cell death in melanoma cell lines [70, 71]. The importance of MEK in regulating the response of the MAPK signaling cascade has been investigated by Mansour *et al.* [72], who have reported that expression of genetically mutated constitutively active MEK is sufficient to cause cellular transformation. In addition to Pho1 and MEK, the SOSA results reveal two other influential species, namely MEK-P-Pho2 and MEK-P, which influence the duration and strength of ERK-PP activity, respectively. Both are key reactant species in the MEK dephosphorylation step of the MAPK signaling cascade that leads to MEK inactivation. It has been reported by Orth *et al.* [73] that the prevention of MEK activation may contribute to eventual apoptosis.

A number of investigators have recently studied sensitivity properties of the MAPK signaling cascade by using derivative-based approaches. In particular, Mayawala *et al.* [50] have considered three response characteristics associated with ERK-PP ac-

tivity, namely decay time, peak time, and amplitude gain, and concluded that there is no reaction that influences the peak time. We can relate decay time and amplitude to the duration and strength of ERK-PP activity, respectively. Mayawala *et al.* concluded that the decay time (duration) is very sensitive to phosphatase reactions at the MEK level and that these reactions do not significantly influence the amplitude gain (strength). These conclusions are in agreement with our results obtained by variance-based sensitivity analysis. Mayawala *et al.* have also concluded that the amplitude gain (strength) is most sensitive to phosphatase reactions at the ERK level. Unfortunately, this conclusion is not supported by our study. Recall that, for small fluctuations (i.e., when $\lambda^{\ddagger} = 0.1$), the ROSA results indicate that reaction 19 (ERK-PP-Pho3 $\rightarrow$ ERK-P + Pho3, dephosphorylation of ERK-PP) influences the strength of ERK-PP activity, whereas, for large fluctuations (i.e., when $\lambda^{\ddagger} = 0.4$), reaction 21 (ERK-P-Pho3 $\rightarrow$ ERK + Pho3, dephosphorylation of ERK-P) influences the strength of ERK-PP activity; see Fig. 3.1. However, our results (data not shown) reveal that the influence of reaction 19 accounts for only 13%, for $\lambda^{\ddagger} = 0.1$, of the duration variance, whereas, the influence of reaction 21 accounts for only 18%, for $\lambda^{\ddagger} = 0.4$, of the strength variance.

On the other hand, Liu *et al.* [17] have identified the activation/inactivation of Raf and several reactions associated with the phosphorylation and dephosphorylation of MEK as being very influential in shaping system output. Their conclusions are in agreement with our results obtained by variance-based sensitivity analysis.

Finally, Hornberg *et al.* [16] have concluded that most reactions in the MAPK signaling cascade are not important for influencing the output ERK-PP profile, and

they have noted that Raf inactivation and MEK phosphorylation by Raf are the most influential processes for controlling the duration, integrated response, and amplitude (related to the strength) of ERK-PP activity. Hornberg *et al.* have also noted that the amplitude is also controlled by ERK phosphorylation and dephosphorylation, whereas, the duration and integrated response are further controlled by MEK dephosphorylation. In general, these results are in agreement with our results obtained by variance-based sensitivity analysis, except for the integrated response, in which case MEK dephosphorylation reactions are identified as non-influential by variance-based sensitivity analysis. Moreover, our results clearly indicate that reaction 4 (Raf* + Pho1 $\rightleftharpoons$ Raf*-Pho1) is the most influential reaction and reaction 6 (MEK + Raf* $\rightleftharpoons$ MEK-Raf*) is the second most influential reaction for controlling the duration, integrated response and strength of ERK-PP activity. In addition, they have studied how changes in the concentration of the molecular species may influence the response characteristics. Alternatively, the SOSA results presented in this dissertation show how the system response is affected when the kinetic properties of molecular species are changed, for example, by genetic variations. The Hornberg *et al.* concentration-based sensitivity analysis results indicate the importance of MEK, ERK, Pho1, Pho2, and Pho3 for influencing the duration and integrated response, as well as the importance of ERK and Pho3 for influencing the amplitude, at small fluctuation levels. Moreover, their results indicate the importance of ERK, Pho1, Pho3, and MEK for influencing the amplitude at large fluctuation levels. On the other hand, the SOSA results presented above clearly identify that Pho1 and MEK are two most influential species in shaping the response profile of ERK-PP, which turns out to have a better specificity and be pharmaceutically more relevant.

Finally, Hornberg *et al.* have noted a high correlation between the control of duration and integrated response and very little correlation between the control of duration and amplitude. Our results agree with their second conclusion, but do not fully agree with the first one. Our results indicate that the integrated response has a larger correlation with duration for small factor fluctuations, whereas, it has a larger correlation with strength for large factor fluctuations. In fact, the sensitivity properties of the integrated response could reflect the combined properties of the duration and strength, since Eq. (2.7) in Section 2.3 implies that the integrated response is the product of duration and strength.

Differences between the previous results and the sensitivity analysis results presented in this dissertation are partly due to the fact that various studies use different choices for the input factors and response characteristics. We also believe that another source of discrepancy is the use of derivatives for sensitivity analysis by the previously mentioned methods. But the most serious problem seems to be the fact that these methods do not employ a thermodynamically consistent approach to sensitivity analysis. It has been recently pointed out by Ederer and Gilles [30] that thermodynamic inconsistencies may lead to erroneous sensitivity analysis results, which may in turn lead to misleading biological conclusions. The method presented in this dissertation effectively addresses this important problem.

We should finally point out that experimental evidence may suggest that a particular cellular behavior of interest is controlled by the combined influence of the concentration profiles of two (or more) molecular species (e.g., Raf* and ERK-PP in the MAPK signaling cascade). A simple way to deal with this case is to separately cal-

culate the SESI's $\sigma_j^{(1)}$ and $\sigma_j^{(2)}$, the TESI's $\tau_j^{(1)}$ and $\tau_j^{(2)}$, and the JESI's $\eta_j^{(1)}$ and $\eta_j^{(2)}$, corresponding to the concentration profiles of two molecular species, say 1 and 2, and set $\sigma_j = \max\{\sigma_j^{(1)}, \sigma_j^{(2)}\}$ and $\eta_j = \max\{\eta_j^{(1)}, \eta_j^{(2)}\}$. It is not difficult to see that the rules for interpreting the variance-based sensitivity indices summarized in Table 3.1 can be applied here as well. For example, if $\eta_j \simeq 0$ and $\sigma_j \simeq 0$, then we may say that factor j does not appreciably influence system behavior. On the other hand, if $\eta_j \simeq 0$ and $\sigma_j \not\simeq 0$, we may say that factor j influences system behavior mostly singularly. In this case however we can get additional information about the actual source of influence by considering the individual SESI values. For example, if $\eta_j \simeq 0$ and $\sigma_j \not\simeq 0$, with $\sigma_j^{(1)} \not\simeq 0$ and $\sigma_j^{(2)} \simeq 0$, then we can say that factor j influences system behavior mostly singularly through the activity of only the first molecular species, whereas, if $\sigma_j^{(1)} \not\simeq 0$ and $\sigma_j^{(2)} \not\simeq 0$, then we can say that factor j influences system behavior mostly singularly through the activity of both molecular species.

3.5 Appendix

3.5.1 Monte Carlo Estimation

Let X and Y be two statistically independent random variables with probability density functions $f_X(x)$ and $f_Y(y)$, respectively, and $Z = g(X, Y)$ be another random variable, which is a function of X and Y. The k^{th}-order moment $\mathrm{E}\left[Z^k\right]$ of Z is given by

$$\mathrm{E}\left[Z^k\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[g(x,y)]^k f_X(x) f_Y(y)\, dxdy. \tag{3.41}$$

We can derive a Monte Carlo estimator $\widehat{\mathrm{E}}\left[Z^k\right]$ for this quantity by independently drawing samples $\{x^{(l)},\, l = 1, 2, \ldots, L\}$ of X and $\{y^{(l)}, l = 1, 2, \ldots, L\}$ of Y from the probability density functions $f_X(x)$ and $f_Y(y)$, respectively, and by setting

$$\widehat{\mathrm{E}}\left[Z^k\right] = \frac{1}{L}\sum_{l=1}^{L}\left[g(x^{(l)}, y^{(l)})\right]^k. \tag{3.42}$$

As a special case, we can estimate the mean and variance of Z by

$$\widehat{\mathrm{E}}\left[Z\right] = \frac{1}{L}\sum_{l=1}^{L} g(x^{(l)}, y^{(l)}), \tag{3.43}$$

$$\widehat{\mathrm{Var}}\left[X\right] = \frac{1}{L}\sum_{l=1}^{L} g^2(x^{(l)}, y^{(l)}) - \widehat{\mathrm{E}}^2\left[Z\right], \tag{3.44}$$

since $\mathrm{Var}\left[Z\right] = \mathrm{E}\left[Z^2\right] - \mathrm{E}^2\left[Z\right]$.

To derive a Monte Carlo estimator for the variance $\mathrm{Var}\big[\mathrm{E}\big[g(X,Y) \mid Y\big]\big]$, note that

$$\begin{aligned}
&\mathrm{Var}\big[\mathrm{E}\big[g(X,Y) \mid Y\big]\big] \\
&= \int_{-\infty}^{\infty} \mathrm{E}^2\big[g(X,Y) \mid Y\big] f_Y(y)\, dy - \big\{\mathrm{E}\big[\mathrm{E}\big[g(X,Y) \mid Y\big]\big]\big\}^2 \\
&= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} g(x_1, y) f_X(x_1)\, dx_1\right]\left[\int_{-\infty}^{\infty} g(x_2, y) f_X(x_2)\, dx_2\right] f_Y(y)\, dy - \mathrm{E}^2\big[g(X,Y)\big] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x_1, y) g(x_2, y) f_X(x_1) f_X(x_2) f_Y(y)\, dx_1 dx_2 dy - \mathrm{E}^2\left[Z\right].
\end{aligned} \tag{3.45}$$

This leads to the following Monte Carlo estimator:

$$\widehat{\mathrm{Var}}\big[\mathrm{E}\big[g(X,Y) \mid Y\big]\big] = \frac{1}{L}\sum_{l=1}^{L} g(x_1^{(l)}, y^{(l)}) g(x_2^{(l)}, y^{(l)}) - \widehat{\mathrm{E}}^2\left[Z\right], \tag{3.46}$$

where $\{x_1^{(l)}, l = 1, 2, \ldots, L\}$, $\{x_2^{(l)}, l = 1, 2, \ldots, L\}$ are two sets of samples of X drawn independently from the probability density function $f_X(x)$, and $\{y^{(l)}, l = 1, 2, \ldots, L\}$ is a set of samples of Y drawn independently from the probability density function $f_Y(y)$.

3.5.2 Proof of Conditions C.1–C.7

In this subsection, we show that conditions C.1-C.7 are indeed satisfied by the Monte Carlo variance estimators introduced in Section 3.2.

Condition C.1 is a direct consequence of the fact that

$$\widehat{\mathrm{Var}}_j\left[R(\mathbf{U})\right] = \frac{1}{4L}\Big\{\sum_{l=1}^{L}\left[R(\mathbf{u}^{(l)}) - R(\mathbf{u}^{(L+l)})\right]^2 + \sum_{l=1}^{L}\left[R(\mathbf{u}_j^{(l)}) - R(\mathbf{u}_{(j)}^{(l)})\right]^2\Big\}, \quad (3.47)$$

and

$$\widehat{\mathrm{Var}}_{jj'}\left[R(\mathbf{U})\right] = \frac{1}{4L}\Big\{\sum_{l=1}^{L}\left[R(\mathbf{u}_j^{(l)}) - R(\mathbf{u}_{(j)}^{(l)})\right]^2 + \sum_{l=1}^{L}\left[R(\mathbf{u}_{j'}^{(l)}) - R(\mathbf{u}_{(j')}^{(l)})\right]^2\Big\}. \quad (3.48)$$

Conditions C.2, C.3, C.5, and C.6 can be shown from Eq. (3.47) and Eq. (3.48) above, the fact that

$$\widehat{\mathrm{Var}}\left[\mathrm{E}\left[R(\mathbf{U}) \mid U_j\right]\right] = \frac{1}{2L}\Big\{\sum_{l=1}^{L}\left[R(\mathbf{u}^{(l)}) - R(\mathbf{u}_{(j)}^{(l)})\right]\left[R(\mathbf{u}_j^{(l)}) - R(\mathbf{u}^{(L+l)})\right]\Big\}, \quad (3.49)$$

$$\widehat{\mathrm{Var}}\left[\mathrm{E}\left[R(\mathbf{U}) \mid \mathbf{U}_{(j)}\right]\right] = \frac{1}{2L}\Big\{\sum_{l=1}^{L}\left[R(\mathbf{u}^{(l)}) - R(\mathbf{u}_j^{(l)})\right]\left[R(\mathbf{u}_{(j)}^{(l)}) - R(\mathbf{u}^{(L+l)})\right]\Big\}, \quad (3.50)$$

$$\widehat{\mathrm{Var}}\left[\mathrm{E}\left[R(\mathbf{U}) \mid U_j, U_{j'}\right]\right] = \frac{1}{2L}\Big\{\sum_{l=1}^{L}\left[R(\mathbf{u}_j^{(l)}) - R(\mathbf{u}_{j'}^{(l)})\right]\left[R(\mathbf{u}_{(j')}^{(l)}) - R(\mathbf{u}_{(j)}^{(l)})\right]\Big\}, \quad (3.51)$$

and the facts that fixing U_j implies that $\mathbf{u}^{(l)} = \mathbf{u}_{(j)}^{(l)}$ and $\mathbf{u}^{(L+l)} = \mathbf{u}_j^{(l)}$, fixing $\mathbf{U}_{(j)}$ implies that $\mathbf{u}^{(l)} = \mathbf{u}_j^{(l)}$ and $\mathbf{u}^{(L+l)} = \mathbf{u}_{(j)}^{(l)}$, whereas, fixing $\mathbf{U}_{(j,j')}$ implies that $\mathbf{u}_j^{(l)} = \mathbf{u}_{(j')}^{(l)}$ and $\mathbf{u}_{(j)}^{(l)} = \mathbf{u}_{j'}^{(l)}$. To show Condition C.4, we define

$$h_1(l) := R(\mathbf{u}^{(l)})$$

$$h_2(l) := R(\mathbf{u}^{(L+l)})$$

$$h_3(l) := R(\mathbf{u}_j^{(l)})$$

$$h_4(l) := R(\mathbf{u}_{(j)}^{(l)}).$$

Then, by using some straightforward algebra, we can show that

$$\{[h_1(l) + h_2(l)] - [h_3(l) + h_4(l)]\}^2 \geq 0$$

implies

$$\begin{aligned}
&\frac{1}{4}\left[h_1^2(l) + h_2^2(l) + h_3^2(l) + h_4^2(l)\right] - \frac{1}{2}\left[h_1(l)h_2(l) + h_3(l)h_4(l)\right] \\
&\qquad \geq \frac{1}{2}\left[h_1(l)h_3(l) + h_2(l)h_4(l)\right] - \frac{1}{2}\left[h_1(l)h_2(l) + h_3(l)h_4(l)\right] \\
&\qquad\qquad + \frac{1}{2}\left[h_1(l)h_4(l) + h_2(l)h_3(l)\right] - \frac{1}{2}\left[h_1(l)h_2(l) + h_3(l)h_4(l)\right], \qquad (3.52)
\end{aligned}$$

which in turn implies that $\widehat{\mathrm{Var}}_j\big[R(\mathbf{U})\big] \geq \widehat{\mathrm{Var}}\big[\mathrm{E}\big[R(\mathbf{U}) \mid U_j\big]\big] + \widehat{\mathrm{Var}}\big[\mathrm{E}\big[R(\mathbf{U}) \mid \mathbf{U}_{(j)}\big]\big]$. Condition C.7 can be shown similarly.

Chapter 4

Approximation Techniques

4.1 Motivation and Background

In the previous two chapters, we have proposed a probabilistic sensitivity analysis approach for biochemical reaction systems that uses the standard chemical potentials of the activated complexes of the underlying reactions and the standard chemical potentials of molecular species as the biochemical factors of interest, and propagates factor uncertainty to a given system response in a thermodynamically consistent manner. Moreover, we have adopted a formal statistical approach to sensitivity analysis, known as variance-based sensitivity analysis [5, 23, 25, 27], which uses a set of indices to quantify the contribution of individual biochemical factors to the variance of the system response.

Unfortunately, it is not in general possible to analytically evaluate variance-based sensitivity indices. As a consequence, these indices are estimated by Monte Carlo sampling [23,27,31,32], which requires evaluation of the system response at each sample.

A major drawback of this approach is its slow rate of convergence. As a matter of fact, the error produced by a naive Monte Carlo estimation approach decreases with an error rate of $O(1/\sqrt{L})$, where L is the number of Monte Carlo samples used [33]. Hence, accurate estimation of the sensitivity indices requires a large number of Monte Carlo samples and, therefore, a large number of system response evaluations. This makes Monte Carlo estimation of variance-based sensitivity indices computationally very expensive, especially in the case of biochemical reaction systems comprised of a large number of reactions and molecular species.

To reduce the computational load of Monte Carlo estimation, it is imperative that we develop techniques which can produce sufficiently accurate estimates of the sensitivity indices in a fraction of the time required by Monte Carlo sampling. In this chapter, we present four such techniques and apply them to the MAPK signaling cascade. We use analytical derivations and sensitivity analysis results, generated by the four methods, to clarify the relative merits of each approximation technique and produce useful insights on when these techniques can be used for sensitivity analysis of biochemical reaction systems.

The first technique is based on a second-order Taylor series expansion, which is an extension of the first-order derivative-based approach for variance-based sensitivity analysis discussed in [5,23,27,34] by including second-order derivative terms. The other three approximation techniques are based on the high-dimensional model representation (HDMR) schemes developed by H. Rabitz and his coworkers [35–37]. Two types of HDMR's have been proposed in the literature: *finite-difference* (or *cut*) HDMR (FD-HDMR) [35–37, 74] and *analysis-of-variance* HDMR (ANOVA-HDMR) [35–37]. An

attractive property of these schemes, which is important in the context of sensitivity analysis, is the fact that the basis functions provide an *exact* hierarchical decomposition of the system response. Each term in the decomposition encapsulates the contribution of an input biochemical factor to the system response singularly or jointly with other factors. It has been argued in [35–37] that, for most physical systems of interest, the response function can be well approximated by a truncated HDMR that uses only low-order basis functions. As a consequence, it is natural to approximate the response function of a biochemical reaction system with a truncated HDMR that includes only first- and second-order basis functions.

A number of alternative approximation techniques for variance-based sensitivity analysis have been proposed in the literature [75–78]. In these techniques, the original response function is approximated by a surrogate function that can be evaluated much more efficiently, and the sensitivity indices are then estimated by Monte Carlo sampling based on the surrogate function. The computational cost of these techniques includes two parts. The first part is on building the surrogate function by a sufficient number of training points, where the original response function values must be evaluated. The second part is on estimating the sensitivity indices by Monte Carlo sampling based on the surrogate function. Computational savings are due to less time required for computing the surrogate function at each Monte Carlo sample than using the original response function (whose evaluation requires solving a system of ordinary differential equations). To estimate the variance-based sensitivity indices, a large number of samples are required to sufficiently reduce the Monte Carlo estimation error, especially when an approximation error already exists by using a surrogate function to replace

the original response function. For large biochemical reaction systems, the entire estimation process can still be very slow, even when evaluation of the surrogate function at each sample point is fast. By contrast, the main advantage of the techniques discussed in this chapter is that they lead to analytical formulas for the sensitivity indices in terms of certain parameters, and thus avoid Monte Carlo estimation of sensitivity indices. As a consequence, the computational cost for calculating the variance-based sensitivity indices reduces to that of estimating the underlying parameters, which leads to appreciable computational savings over the techniques proposed in [75–78].

4.2 Second-Order Sensitivity Indices

In Chapter 2, we used variance-based sensitivity analysis technique to assess how uncertainty in the rate constants of a biochemical reaction system can affect the system response. In most practical situations, it is difficult to evaluate the high-order terms (≥ 3) in the response variance decomposition scheme given by Eq. (3.1). Although these terms are usually negligible at low to moderate levels of biochemical factor uncertainty, they may take substantial values at high levels [32]. Unfortunately, it is difficult to deal in practice with high-order variance terms. For this reason, it is quite convenient to base our sensitivity analysis effort only on the first- and second-order terms V_j and $V_{jj'}$. Then, instead of using the total system response variance V, we base our sensitivity analysis on its second-order portion $V^{(2)}$, given by

$$V^{(2)} = \sum_{j=1}^{J} V_j + \sum_{j=1}^{J-1} \sum_{j'=j+1}^{J} V_{jj'}. \tag{4.1}$$

By using the probabilistic model given by Eq. (2.21) and the variance decomposition

scheme in Eq. (4.1), we can develop a powerful (second-order) methodology for sensitivity analysis of biochemical reaction systems, similar to the one discussed in Chapter 3, which was based on the total response variance V. The method requires evaluation of two indices, namely the (second-order) *single-effect sensitivity index* (SESI) $\sigma_j^{(2)}$, defined by

$$\sigma_j^{(2)} := \frac{V_j}{V^{(2)}}, \tag{4.2}$$

and the (second-order) *joint-effect sensitivity index* (JESI) $\eta_j^{(2)}$, defined by

$$\eta_j^{(2)} := \frac{T_j}{V^{(2)}}, \tag{4.3}$$

where

$$T_j := \sum_{j=1, j' \neq j}^{J} V_{jj'}. \tag{4.4}$$

Clearly, $\sigma_j^{(2)}$ quantifies the fractional singular contribution of the j^{th} biochemical factor to the second-order portion $V^{(2)}$ of the total response variance V, whereas, $\eta_j^{(2)}$ quantifies the fractional contribution of the j^{th} biochemical factor to $V^{(2)}$ jointly with another factor. It turns out that, if $\sigma_j^{(2)} = \eta_j^{(2)} = 0$, then we can conclude that factor j does not influence the system response singularly or jointly with another factor (although, it may influence the system response jointly with two or more factors). On the other hand, if $\sigma_j^{(2)} > 0$ and $\eta_j^{(2)} = 0$, then we can conclude that factor j influences the system response singularly but not jointly with another factor. Moreover, if $\sigma_j^{(2)} = 0$ and $\eta_j^{(2)} > 0$, we can conclude that factor j does not influence the system response singularly but it does so jointly with some other factor, whereas, if $\sigma_j^{(2)} > 0$ and $\eta_j^{(2)} > 0$, we can conclude that factor j influences the system response both singularly and jointly with some other factor. In practice, we can set a small threshold θ to determine whether

$\sigma_j^{(2)}$ and $\eta_j^{(2)}$ are sufficiently larger than zero.

A straightforward technique for approximating the SESI and JESI values is based on a Monte Carlo Latin hypercube sampling approach, whose details can be found in Section 3.2 (see also [23, 31, 32]). This approach can be used to provide estimates $\widehat{\sigma}_j^{(2)}$ and $\widehat{\eta}_j^{(2)}$ of the second-order SESI's and JESI's by using $2L(J+1)$ system evaluations [by integrating the system of N ordinary differential equations given by Eq. (2.2)], where L is the number of Latin hypercube samples used and J is the number of biochemical factors considered in the analysis. We refer to $\widehat{\sigma}_j^{(2)}$ and $\widehat{\eta}_j^{(2)}$ as the (second-order) SESI's and JESI's obtained by *Monte Carlo* (MC) estimation. This method is computationally expensive, since a large number L of Latin hypercube samples is required to obtain sufficiently accurate estimates of the sensitivity indices.

In this chapter, we present several other methods for approximating the indices $\sigma_j^{(2)}$ and $\eta_j^{(2)}$ associated with the second-order variance-based sensitivity analysis. We first review a number of multivariate representation schemes for the response function of a biochemical reaction system that can be used to analytically map the complex relationship between the biochemical factors and the system response. We then discuss how to use these schemes in order to approximate $\sigma_j^{(2)}$ and $\eta_j^{(2)}$. We also present details regarding the numerical implementation of the resulting approximation techniques.

4.3 Response Function Representation

For ease of presentation, we will base our discussion on a response function $R(\mathbf{u}) = R(u_1, u_2, u_3)$ that depends only on three factors of interest $\mathbf{u} = \{u_1, u_2, u_3\}$. Although

extension to the case of J biochemical factors is straightforward, the required notation is cumbersome and makes key steps difficult to follow. For this reason, we use a trivariate response function to derive key equations and state the general form of these equations without proof.

4.3.1 TSMR

If the response function R is continuously differentiable in a neighborhood of $\mathbf{u} = 0$, then its Taylor series expansion about $\mathbf{0}$ is given by

$$\begin{aligned} R(u_1, u_2, u_3) = r_0 &+ r_1(u_1) + r_2(u_2) + r_3(u_3) + r_{12}(u_1, u_2) \\ &+ r_{13}(u_1, u_3) + r_{23}(u_2, u_3) + r_{123}(u_1, u_2, u_3), \end{aligned} \tag{4.5}$$

where

$$\begin{aligned} r_0 &:= R(\mathbf{0}) \\ r_1(u_1) &:= \sum_{m=1}^{\infty} \frac{1}{m!} \frac{\partial^m R(\mathbf{0})}{\partial u_1^m} u_1^m \\ r_2(u_2) &:= \sum_{m=1}^{\infty} \frac{1}{m!} \frac{\partial^m R(\mathbf{0})}{\partial u_2^m} u_2^m \\ r_3(u_3) &:= \sum_{m=1}^{\infty} \frac{1}{m!} \frac{\partial^m R(\mathbf{0})}{\partial u_3^m} u_3^m \end{aligned}$$

$$r_{12}(u_1,u_2) := \sum_{m_1=1}^{\infty}\sum_{m_2=1}^{\infty} \frac{1}{m_1!\,m_2!}\frac{\partial^{m_1+m_2}R(\mathbf{0})}{\partial u_1^{m_1}\partial u_2^{m_2}}\, u_1^{m_1}u_2^{m_2}$$

$$r_{13}(u_1,u_3) := \sum_{m_1=1}^{\infty}\sum_{m_3=1}^{\infty} \frac{1}{m_1!\,m_3!}\frac{\partial^{m_1+m_3}R(\mathbf{0})}{\partial u_1^{m_1}\partial u_3^{m_3}}\, u_1^{m_1}u_3^{m_3}$$

$$r_{23}(u_2,u_3) := \sum_{m_2=1}^{\infty}\sum_{m_3=1}^{\infty} \frac{1}{m_2!\,m_3!}\frac{\partial^{m_2+m_3}R(\mathbf{0})}{\partial u_2^{m_2}\partial u_3^{m_3}}\, u_2^{m_2}u_3^{m_3}$$

$$r_{123}(u_1,u_2,u_3) := \sum_{m_1=1}^{\infty}\sum_{m_2=1}^{\infty}\sum_{m_3=1}^{\infty} \frac{1}{m_1!\,m_2!\,m_3!}\frac{\partial^{m_1+m_2+m_3}R(\mathbf{0})}{\partial u_1^{m_1}\partial u_2^{m_2}\partial u_3^{m_3}}\, u_1^{m_1}u_2^{m_2}u_3^{m_3}. \quad (4.6)$$

Clearly, the Taylor series expansion provides a representation of the system response R in terms of functions r, given by Eq. (4.6). We refer to the r's as *basis* functions. Note that r_0 is the value of R at the reference point $\mathbf{0}$. On the other hand, $r_1(u_1)$ summarizes the *singular* contribution of factor u_1 to the value of R, whereas, $r_{12}(u_1,u_2)$ summarizes the *joint* contribution of factors u_1 and u_2. Finally, $r_{123}(u_1,u_2,u_3)$ summarizes the joint contribution of all three factors to the value of R. Similar remarks apply for r_2, r_3, r_{13}, and r_{23}.

Although Eq. (4.6) provides analytical formulas for the basis functions, calculating these functions at a point $\mathbf{u}$ requires knowledge of the partial derivatives of R at the reference point $\mathbf{0}$, as well as evaluation of infinite sums, which is very difficult to do in practice. Note however that any basis function r given by Eq. (4.6) is zero if one of its arguments equals zero. By using this property and Eq. (4.5), we have that

$$R(0,0,0) = r_0$$

$$R(u_1,0,0) = r_0 + r_1(u_1)$$

$$R(0,u_2,0) = r_0 + r_2(u_2)$$

$$R(0,0,u_3) = r_0 + r_3(u_3)$$

$$R(u_1, u_2, 0) = r_0 + r_1(u_1) + r_2(u_2) + r_{12}(u_1, u_2)$$

$$R(u_1, 0, u_3) = r_0 + r_1(u_1) + r_3(u_3) + r_{13}(u_1, u_3)$$

$$R(0, u_2, u_3) = r_0 + r_2(u_2) + r_3(u_3) + r_{23}(u_2, u_3),$$

which results in

$$r_0 = R(0, 0, 0)$$

$$r_1(u_1) = R(u_1, 0, 0) - R(0, 0, 0)$$

$$r_2(u_2) = R(0, u_2, 0) - R(0, 0, 0)$$

$$r_3(u_3) = R(0, 0, u_3) - R(0, 0, 0)$$

$$r_{12}(u_1, u_2) = R(u_1, u_2, 0) - R(u_1, 0, 0) - R(0, u_2, 0) + R(0, 0, 0)$$

$$r_{13}(u_1, u_3) = R(u_1, 0, u_3) - R(u_1, 0, 0) - R(0, 0, u_3) + R(0, 0, 0)$$

$$r_{23}(u_2, u_3) = R(0, u_2, u_3) - R(0, u_2, 0) - R(0, 0, u_3) + R(0, 0, 0)$$

$$\begin{aligned} r_{123}(u_1, u_2, u_3) = {} & R(u_1, u_2, u_3) - R(u_1, u_2, 0) - R(u_1, 0, u_3) - R(0, u_2, u_3) \\ & + R(u_1, 0, 0) + R(0, u_2, 0) + R(0, 0, u_3) - R(0, 0, 0). \end{aligned} \tag{4.7}$$

These formulas provide a method for evaluating the basis functions r at some point $\mathbf{u}$. This can be done by calculating the system response at the corresponding u values suggested by the formulas. For example, evaluation of r_0 requires calculation of the system response at $u_1 = u_2 = u_3 = 0$, whereas, evaluation of $r_1(u_1)$ requires an additional calculation of the system response at $u_1, u_2 = u_3 = 0$. This can be done by solving the system of ordinary differential equations given by Eq. (2.2). We refer to the representation scheme given by Eqs. (4.5) and (4.6) as *Taylor Series Model Representation* (TSMR).

4.3.2 FD-HDMR

We can extend the decomposition scheme given by Eq. (4.5) to the case of J biochemical factors and to functions that are not necessarily continuously differentiable. As a matter of fact, we can represent *any* response function R with J factors $\mathbf{u} = \{u_1, u_2, \ldots, u_J\}$ by

$$R(\mathbf{u}) = r_0 + \sum_{j=1}^{J} \sum_{1 \leq m_1 < \cdots < m_j \leq J} \cdots \sum r_{m_1 m_2 \cdots m_j}(u_{m_1}, u_{m_2}, \ldots, u_{m_j}). \tag{4.8}$$

The only requirement is that we must be able to *uniquely* determine the basis functions r from R. The representation of a multidimensional function R by Eq. (4.8) is known in the literature as *High-Dimensional Model Representation* (HDMR) [35, 36].

A way to guarantee that we can uniquely determine r from the response function R is to consider basis functions that become zero if one of their arguments is zero. In this case, r can be determined by the classical Möbius inversion formula

$$r_{m_1 m_2 \cdots m_j}(u_{m_1}, u_{m_2}, \ldots, u_{m_j}) = \sum_{I' \subseteq I} (-1)^{|I \setminus I'|} R(\mathbf{u}_{I'}), \tag{4.9}$$

which generalizes Eq. (4.7). In this formula, $I = \{m_1, m_2, \ldots, m_j\}$, $A \setminus B$ denotes the set difference between two sets A and B, $|A|$ denotes the number of elements in a set A (by convention, we set $|\emptyset| = 0$), and $\mathbf{u}_{I'}$ is $\mathbf{u}$ with all variables, except the one indexed by I', set to zero.

Eqs. (4.8) and (4.9) express $R(\mathbf{u})$ as a superposition of system response values on lines, planes and hyperplanes passing through the reference point $\mathbf{0}$. For this reason, these equations lead to a system representation scheme known in the literature as cut-HDMR [35–37] or *Finite Difference* (FD) HDMR [74]. We adopt the second terminology here as being more appropriate for characterizing this type of HDMR. Clearly,

the Taylor series expansion is a special case of FD-HDMR, with basis functions given by Eq. (4.7).

4.3.3 ANOVA-HDMR

Let us now assume that we can find invertible differentiable transformations g_j, which we can use to map the biochemical factors u_j into factors $s_j := g_j(u_j)$, that take values between 0 and 1. Let

$$P(s_1, s_2, s_3) := R(g_1^{-1}(s_1), g_2^{-1}(s_2), g_3^{-1}(s_3)). \tag{4.10}$$

The HDMR representation of P is given by

$$\begin{aligned} P(s_1, s_2, s_3) = p_0 &+ p_1(s_1) + p_2(s_2) + p_3(s_3) + p_{12}(s_1, s_2) \\ &+ p_{13}(s_1, s_3) + p_{23}(s_2, s_3) + p_{123}(s_1, s_2, s_3). \end{aligned} \tag{4.11}$$

If we consider basis functions p whose integrals over a *single* variable are equal to zero, then we can readily verify from Eq. (4.11) that

$$\begin{aligned} p_0 &= \int_0^1 \int_0^1 \int_0^1 P(s_1, s_2, s_3) ds_1 ds_2 ds_3 \\ p_1(s_1) &= \int_0^1 \int_0^1 P(s_1, s_2, s_3) ds_2 ds_3 - p_0 \\ p_2(s_2) &= \int_0^1 \int_0^1 P(s_1, s_2, s_3) ds_1 ds_3 - p_0 \\ p_3(s_3) &= \int_0^1 \int_0^1 P(s_1, s_2, s_3) ds_1 ds_2 - p_0 \\ p_{12}(s_1, s_2) &= \int_0^1 P(s_1, s_2, s_3) ds_3 - p_1(s_1) - p_2(s_2) - p_0 \\ p_{13}(s_1, s_3) &= \int_0^1 P(s_1, s_2, s_3) ds_2 - p_1(s_1) - p_3(s_3) - p_0 \end{aligned}$$

$$p_{23}(s_2, s_3) = \int_0^1 P(s_1, s_2, s_3) ds_1 - p_2(s_2) - p_3(s_3) - p_0$$

$$p_{123}(s_1, s_2, s_3) = P(s_1, s_2, s_3) - p_{12}(s_1, s_2) - p_{13}(s_1, s_3) - p_{23}(s_2, s_3)$$
$$- p_1(s_1) - p_2(s_2) - p_3(s_3) - p_0. \tag{4.12}$$

Therefore, we can uniquely determine the basis functions p from P. By setting $s_j = g_j(u_j)$ in Eqs. (4.11) and (4.12), and by employing Eq. (4.10), we obtain

$$R(u_1, u_2, u_3) = \rho_0 + \rho_1(u_1) + \rho_2(u_2) + \rho_3(u_3) + \rho_{12}(u_1, u_2)$$
$$+ \rho_{13}(u_1, u_3) + \rho_{23}(u_2, u_3) + \rho_{123}(u_1, u_2, u_3), \tag{4.13}$$

where

$$\rho_0 := p_0 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R(u_1, u_2, u_3) g_1'(u_1) g_2'(u_2) g_3'(u_3) du_1 du_2 du_3$$

$$\rho_1(u_1) := p_1(g_1(u_1)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R(u_1, u_2, u_3) g_2'(u_2) g_3'(u_3) du_2 du_3 - \rho_0$$

$$\rho_2(u_2) := p_2(g_2(u_2)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R(u_1, u_2, u_3) g_1'(u_1) g_3'(u_3) du_1 du_3 - \rho_0$$

$$\rho_3(u_3) := p_3(g_3(u_3)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R(u_1, u_2, u_3) g_1'(u_1) g_2'(u_2) du_1 du_2 - \rho_0$$

$$\rho_{12}(u_1, u_2) := p_{12}(g_1(u_1), g_2(u_2))$$
$$= \int_{-\infty}^{\infty} R(u_1, u_2, u_3) g_3'(u_3) du_3 - \rho_1(u_1) - \rho_2(u_2) - \rho_0$$

$$\rho_{13}(u_1, u_3) := p_{13}(g_1(u_1), g_3(u_3))$$
$$= \int_{-\infty}^{\infty} R(u_1, u_2, u_3) g_2'(u_2) du_2 - \rho_1(u_1) - \rho_3(u_3) - \rho_0$$

$$
\begin{aligned}
\rho_{23}(u_2,u_3) &:= p_{23}(g_2(u_2),g_3(u_3)) \\
&= \int_{-\infty}^{\infty} R(u_1,u_2,u_3)g_1'(u_1)du_1 - \rho_2(u_2) - \rho_3(u_3) - \rho_0 \\
\rho_{123}(u_1,u_2,u_3) &:= p_{123}(g_1(u_1),g_2(u_2),g_3(u_3)) \\
&= R(u_1,u_2,u_3) - \rho_{12}(u_1,u_2) - \rho_{13}(u_1,u_3) - \rho_{23}(u_2,u_3) \\
&\quad -\rho_1(u_1) - \rho_2(u_2) - \rho_3(u_3) - \rho_0, \qquad (4.14)
\end{aligned}
$$

with g' being the first-order derivative of g. For reasons to be explained in Subsection 4.4.4, the representation of a response function R by Eqs. (4.13) and (4.14) is referred to in the literature as *Analysis-of-Variance* (ANOVA) HDMR [26, 35–37, 74]. Note that the basis functions ρ satisfy the following orthogonality conditions:

$$
\begin{aligned}
&\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \rho_{j_1,\ldots,j_k}(u_{j_1},\ldots,u_{j_k})g_1'(u_1)\cdots g_J'(u_J)du_1\cdots du_J = 0, \\
&\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \rho_{j_1,\ldots,j_k}(u_{j_1},\ldots,u_{j_k})\rho_{j_1',\ldots,j_{k'}'}(u_{j_1'},\ldots,u_{j_{k'}'})g_1'(u_1)\cdots g_J'(u_J)du_1\cdots du_J = 0, \\
&\qquad\qquad (j_1,\ldots,j_k) \neq (j_1',\ldots,j_{k'}'), \qquad (4.15)
\end{aligned}
$$

provided that $\int_{-\infty}^{\infty} g_j'(u_j)du_j = 1$, for $j = 1, 2, \ldots, J$.

4.4 Approximations of Response Variances

In most applications of interest, it is very difficult to directly evaluate the variance-based sensitivity indices by means of Eqs. (4.1)–(4.4), due to the complexity of the response function R. We can address this problem by replacing the response function with a simpler function $\widehat{R}(u_1,u_2,\ldots,u_J)$ that will allow us to approximate the response

variances given by

$$\begin{aligned} V_j &:= \mathrm{Var}\big[\mathrm{E}\big[R(\mathbf{U}) \mid U_j\big]\big], \\ V_{jj'} &:= \mathrm{Var}\big[\mathrm{E}\big[R(\mathbf{U}) \mid U_j, U_{j'}\big]\big] - \mathrm{Var}\big[\mathrm{E}\big[R(\mathbf{U}) \mid U_j\big]\big] - \mathrm{Var}\big[\mathrm{E}\big[R(\mathbf{U}) \mid U_{j'}\big]\big], \end{aligned} \tag{4.16}$$

In the following, we discuss various approximations obtained by employing the previously discussed representation schemes.

4.4.1 Derivative Approximation

As we mentioned in Subsection 4.3.1, the two main problems associated with the basis functions of TSMR, given by Eq. (4.6), is the need to calculate high-order partial derivatives of the response function and to evaluate infinite sums. To address these problems, we can approximate the basis functions by assuming that the response function is sufficiently smooth in a neighborhood around $\mathbf{0}$ so that partial derivatives of orders greater than two are negligible. In this case, we can approximate the response function $R(u_1, u_2, u_3)$ by

$$\widehat{R}(u_1, u_2, u_3) = \widehat{r}_0 + \widehat{r}_1(u_1) + \widehat{r}_2(u_2) + \widehat{r}_3(u_3) + \widehat{r}_{12}(u_1, u_2) + \widehat{r}_{13}(u_1, u_3) + \widehat{r}_{23}(u_2, u_3), \tag{4.17}$$

where

$$\begin{aligned} \widehat{r}_0 &:= R(\mathbf{0}) \\ \widehat{r}_1(u_1) &:= \frac{\partial R(\mathbf{0})}{\partial u_1}\, u_1 + \frac{1}{2}\frac{\partial^2 R(\mathbf{0})}{\partial u_1^2}\, u_1^2 \\ \widehat{r}_2(u_2) &:= \frac{\partial R(\mathbf{0})}{\partial u_2}\, u_2 + \frac{1}{2}\frac{\partial^2 R(\mathbf{0})}{\partial u_2^2}\, u_2^2 \\ \widehat{r}_3(u_3) &:= \frac{\partial R(\mathbf{0})}{\partial u_3}\, u_3 + \frac{1}{2}\frac{\partial^2 R(\mathbf{0})}{\partial u_3^2}\, u_3^2 \end{aligned}$$

$$\widehat{r}_{12}(u_1, u_2) := \frac{\partial^2 R(\mathbf{0})}{\partial u_1 \partial u_2}\, u_1 u_2$$

$$\widehat{r}_{13}(u_1, u_3) := \frac{\partial^2 R(\mathbf{0})}{\partial u_1 \partial u_3}\, u_1 u_3$$

$$\widehat{r}_{23}(u_2, u_3) := \frac{\partial^2 R(\mathbf{0})}{\partial u_2 \partial u_3}\, u_2 u_3\,, \tag{4.18}$$

since $r_{123}(u_1, u_2, u_3) = 0$ in this case. By employing the statistical independence of U_1, U_2, and U_3, we can show that the variances associated with the approximate response function $\widehat{R}(u_1, u_2, u_3)$ satisfy:

$$\begin{aligned}
\widehat{V} &= d_1^2\lambda_1^2 + d_2^2\lambda_2^2 + d_3^2\lambda_3^2 + \frac{1}{2}\, d_{11}^2\lambda_1^4 + \frac{1}{2}\, d_{22}^2\lambda_2^4 + \frac{1}{2}\, d_{33}^2\lambda_3^4 \\
&\quad + d_{12}^2\lambda_1^2\lambda_2^2 + d_{13}^2\lambda_1^2\lambda_3^2 + d_{23}^2\lambda_2^2\lambda_3^2 \\
\widehat{V}_1 &= d_1^2\lambda_1^2 + \frac{1}{2}\, d_{11}^2\lambda_1^4 \\
\widehat{V}_2 &= d_2^2\lambda_2^2 + \frac{1}{2}\, d_{22}^2\lambda_2^4 \\
\widehat{V}_3 &= d_3^2\lambda_3^2 + \frac{1}{2}\, d_{33}^2\lambda_3^4 \\
\widehat{V}_{12} &= d_{12}^2\lambda_1^2\lambda_2^2 \\
\widehat{V}_{13} &= d_{13}^2\lambda_1^2\lambda_3^2 \\
\widehat{V}_{23} &= d_{23}^2\lambda_2^2\lambda_3^2\,,
\end{aligned} \tag{4.19}$$

where d_j is the first-order partial derivative of R with respect to u_j at $\mathbf{0}$ and $d_{jj'}$ is the second-order partial derivative of R with respect to u_j and $u_{j'}$ at $\mathbf{0}$. To show Eq. (4.19), we have used the fact that U_j follows a Gaussian distribution with zero mean and standard deviation λ_j, which implies $\mathrm{E}\big[U_j^3\big] = 0$, $\mathrm{E}\big[U_j^4\big] = 3\lambda_j^4$. As a consequence of Eqs. (4.1)–(4.4), and Eq. (4.19), we obtain the following approximations to the (second-

order) SESI's and JESI's (expressed for the general case of J biochemical factors):

$$\boxed{\begin{aligned} \widehat{\sigma}_j^{(2)} &= \frac{\widehat{V}_j}{\widehat{V}}, \qquad \widehat{\eta}_j^{(2)} = \frac{\widehat{T}_j}{\widehat{V}} \\ \widehat{V}_j &= \lambda_j^2 d_j^2 + \frac{1}{2}\lambda_j^4 d_{jj}^2 \\ \widehat{V}_{jj'} &= \lambda_j^2 \lambda_{j'}^2 d_{jj'}^2 \\ \widehat{T}_j &= \sum_{j'=1, j' \neq j}^{J} \widehat{V}_{jj'} \\ \widehat{V} &= \sum_{j=1}^{J} \widehat{V}_j + \sum_{j=1}^{J-1} \sum_{j'=j+1}^{J} \widehat{V}_{jj'} \end{aligned}} \tag{4.20}$$

We respectively refer to $\widehat{\sigma}_j^{(2)}$ and $\widehat{\eta}_j^{(2)}$, given by Eq. (4.20), as the (second-order) SESI's and JESI's obtained by *Derivative Approximation* (DA).

4.4.2 Polynomial Approximation of FD-HDMR

We may obtain a better approximation of the sensitivity indices $\sigma_j^{(2)}$ and $\eta_j^{(2)}$ by assuming that the response function is sufficiently smooth in a neighborhood around $\mathbf{0}$ so that partial derivatives of orders greater than two with respect to one variable and partial derivatives that involve more than two variables are negligible. In this case, we can approximate the response function $R(u_1, u_2, u_3)$ by

$$\widehat{R}(u_1, u_2, u_3) = \widehat{r}_0 + \widehat{r}_1(u_1) + \widehat{r}_2(u_2) + \widehat{r}_3(u_3) + \widehat{r}_{12}(u_1, u_2) + \widehat{r}_{13}(u_1, u_3) + \widehat{r}_{23}(u_2, u_3), \tag{4.21}$$

where

$$\widehat{r}_0 := R(\mathbf{0})$$

$$\widehat{r}_1(u_1) := \frac{\partial R(\mathbf{0})}{\partial u_1} u_1 + \frac{1}{2}\frac{\partial^2 R(\mathbf{0})}{\partial u_1^2} u_1^2$$

$$\widehat{r}_2(u_2) := \frac{\partial R(\mathbf{0})}{\partial u_2} u_2 + \frac{1}{2}\frac{\partial^2 R(\mathbf{0})}{\partial u_2^2} u_2^2$$

$$
\begin{aligned}
\widehat{r}_3(u_3) &:= \frac{\partial R(\mathbf{0})}{\partial u_3}\, u_3 + \frac{1}{2}\frac{\partial^2 R(\mathbf{0})}{\partial u_3^2}\, u_3^2 \\
\widehat{r}_{12}(u_1,u_2) &:= \frac{\partial^2 R(\mathbf{0})}{\partial u_1 \partial u_2}\, u_1 u_2 + \frac{1}{2}\frac{\partial^3 R(\mathbf{0})}{\partial u_1^2 \partial u_2}\, u_1^2 u_2 + \frac{1}{2}\frac{\partial^3 R(\mathbf{0})}{\partial u_1 \partial u_2^2}\, u_1 u_2^2 + \frac{1}{4}\frac{\partial^4 R(\mathbf{0})}{\partial u_1^2 \partial u_2^2}\, u_1^2 u_2^2 \\
\widehat{r}_{13}(u_1,u_3) &:= \frac{\partial^2 R(\mathbf{0})}{\partial u_1 \partial u_3}\, u_1 u_3 + \frac{1}{2}\frac{\partial^3 R(\mathbf{0})}{\partial u_1^2 \partial u_3}\, u_1^2 u_3 + \frac{1}{2}\frac{\partial^3 R(\mathbf{0})}{\partial u_1 \partial u_3^2}\, u_1 u_3^2 + \frac{1}{4}\frac{\partial^4 R(\mathbf{0})}{\partial u_1^2 \partial u_3^2}\, u_1^2 u_3^2 \\
\widehat{r}_{23}(u_2,u_3) &:= \frac{\partial^2 R(\mathbf{0})}{\partial u_2 \partial u_3}\, u_2 u_3 + \frac{1}{2}\frac{\partial^3 R(\mathbf{0})}{\partial u_2^2 \partial u_3}\, u_2^2 u_3 + \frac{1}{2}\frac{\partial^3 R(\mathbf{0})}{\partial u_2 \partial u_3^2}\, u_2 u_3^2 + \frac{1}{4}\frac{\partial^4 R(\mathbf{0})}{\partial u_2^2 \partial u_3^2}\, u_2^2 u_3^2.
\end{aligned}
\tag{4.22}
$$

Due to difficulties in numerically evaluating high-order derivatives with sufficient accuracy, we may not be able to use Eq. (4.22) to derive sufficiently good DA approximations of the sensitivity indices. However, this equation motivates us to set

$$
\begin{aligned}
&\widehat{r}_j(u_j) = \alpha_{j,1} u_j + \alpha_{j,2}\, u_j^2, \\
&\widehat{r}_{jj'}(u_j,u_{j'}) = \alpha_{jj',1}\, u_j u_{j'} + \alpha_{jj',2}\, u_j^2 u_{j'} + \alpha_{jj',3}\, u_j u_{j'}^2 + \alpha_{jj',4}\, u_j^2 u_{j'}^2,
\end{aligned}
\tag{4.23}
$$

where the α's are parameters whose values must be appropriately determined so that $\widehat{R}$, given by Eqs. (4.21) and (4.23), sufficiently approximates the response function R. We will be discussing a practical method to address this problem in Subsection 4.5.2 of this chapter.

Clearly, the previous approach is based on approximating the first- and second-order basis functions associated with the FD-HDMR given by Eqs. (4.8) and (4.9) with the polynomials given by Eq. (4.23). If $\widehat{R}(\mathbf{u})$ is sufficiently close to $R(\mathbf{u})$ in a neighborhood around $\mathbf{0}$, then the partial derivatives of R associated with Eq. (4.22) can be obtained from the parameters α. Note that the approximating basis functions $\widehat{r}$ given by Eq. (4.23) satisfy the necessary condition of becoming zero if one of their arguments equals zero.

As a consequence of Eqs. (4.1)–(4.4), by employing the statistical independence of the U_j's, and by using the fact that U_j follows a Gaussian distribution with zero mean and standard deviation λ_j, in which case $\mathrm{E}\big[U_j^3\big] = 0$ and $\mathrm{E}\big[U_j^4\big] = 3\lambda_j^4$, we obtain the following approximations to the (second-order) SESI's and JESI's:

$$
\begin{gathered}
\widehat{\sigma}_j^{(2)} = \frac{\widehat{V}_j}{\widehat{V}}, \qquad \widehat{\eta}_j^{(2)} = \frac{\widehat{T}_j}{\widehat{V}} \\
\widehat{V}_j = \lambda_j^2\alpha_{j,1}^2 + 2\lambda_j^4\alpha_{j,2}^2 + 2\lambda_j^2\alpha_{j,1}\left(\sum_{m=1}^{j-1}\lambda_m^2\alpha_{mj,2} + \sum_{m=j+1}^{J}\lambda_m^2\alpha_{jm,3}\right) \\
+ \lambda_j^2\left(\sum_{m=1}^{j-1}\lambda_m^2\alpha_{mj,2} + \sum_{m=j+1}^{J}\lambda_m^2\alpha_{jm,3}\right)^2 \\
+ 4\lambda_j^4\alpha_{j,2}\left(\sum_{m=1}^{j-1}\lambda_m^2\alpha_{mj,4} + \sum_{m=j+1}^{J}\lambda_m^2\alpha_{jm,4}\right) \\
+ 2\lambda_j^4\left(\sum_{m=1}^{j-1}\lambda_m^2\alpha_{mj,4} + \sum_{m=j+1}^{J}\lambda_m^2\alpha_{jm,4}\right)^2 \\
\widehat{V}_{jj'} = \lambda_j^2\lambda_{j'}^2\alpha_{jj',1}^2 + 2\lambda_j^4\lambda_{j'}^2\alpha_{jj',2}^2 + 2\lambda_j^2\lambda_{j'}^4\alpha_{jj',3}^2 + 4\lambda_j^4\lambda_{j'}^4\alpha_{jj',4}^2 \\
\widehat{T}_j = \sum_{j'=1,j'\neq j}^{J}\widehat{V}_{jj'} \\
\widehat{V} = \sum_{j=1}^{J}\widehat{V}_j + \sum_{j=1}^{J-1}\sum_{j'=j+1}^{J}\widehat{V}_{jj'}
\end{gathered}
\tag{4.24}
$$

Note that Eq. (4.24) is a special case of Equations 35 and 36 in [79]. We respectively refer to $\widehat{\sigma}_j^{(2)}$ and $\widehat{\eta}_j^{(2)}$, given by Eq. (4.24), as the (second-order) SESI's and JESI's obtained by *Polynomial Approximation* (PA) of the FD-HDMR.

4.4.3 Gauss-Hermite Integration of FD-HDMR

We can derive another approximation of the sensitivity indices by assuming that the partial derivatives of the response function in a neighborhood of $\mathbf{0}$ that involve more than two factors are negligible. In this case, we can approximate the response function $R(u_1, u_2, u_3)$ by

$$\widehat{R}(u_1, u_2, u_3) = r_0 + r_1(u_1) + r_2(u_2) + r_3(u_3) + r_{12}(u_1, u_2) + r_{13}(u_1, u_3) + r_{23}(u_2, u_3), \quad (4.25)$$

where

$$\begin{aligned}
r_0 &= R(\mathbf{0}) \\
r_1(u_1) &= \sum_{m=1}^{\infty} \frac{1}{m!} \frac{\partial^m R(\mathbf{0})}{\partial u_1^m} \, u_1^m \\
r_2(u_2) &= \sum_{m=1}^{\infty} \frac{1}{m!} \frac{\partial^m R(\mathbf{0})}{\partial u_2^m} \, u_2^m \\
r_3(u_3) &= \sum_{m=1}^{\infty} \frac{1}{m!} \frac{\partial^m R(\mathbf{0})}{\partial u_3^m} \, u_3^m \\
r_{12}(u_1, u_2) &= \sum_{m_1=1}^{\infty} \sum_{m_2=1}^{\infty} \frac{1}{m_1! m_2!} \frac{\partial^{m_1+m_2} R(\mathbf{0})}{\partial u_1^{m_1} \partial u_2^{m_2}} \, u_1^{m_1} u_2^{m_2} \\
r_{13}(u_1, u_3) &= \sum_{m_1=1}^{\infty} \sum_{m_3=1}^{\infty} \frac{1}{m_1! m_3!} \frac{\partial^{m_1+m_3} R(\mathbf{0})}{\partial u_1^{m_1} \partial u_3^{m_3}} \, u_1^{m_1} u_3^{m_3} \\
r_{23}(u_2, u_3) &= \sum_{m_2=1}^{\infty} \sum_{m_3=1}^{\infty} \frac{1}{m_2! m_3!} \frac{\partial^{m_2+m_3} R(\mathbf{0})}{\partial u_2^{m_2} \partial u_3^{m_3}} \, u_2^{m_2} u_3^{m_3}, \qquad (4.26)
\end{aligned}$$

since $r_{123}(u_1, u_2, u_3) = 0$ in this case. We expect that this approximation will be more accurate than the one considered in Eqs. (4.21) and (4.22), since the first- and second-order basis functions are exactly the same as the corresponding basis functions given by Eq. (4.6). Note that we can obtain the approximation given by Eq. (4.25) by simply truncating the third- and higher-order terms in the FD-HDMR of the response func-

tion R, given by Eqs. (4.8) and (4.9), without making any reference to the derivatives of R.

Since the basis functions r given by Eq. (4.26) become zero if one of their arguments is zero, we can relate them to the system response R by means of Eq. (4.7). As a consequence of Eqs. (4.7) and (4.25), we obtain the following decomposition for $\widehat{R}$ (expressed for the general case of J biochemical factors):

$$\widehat{R}(\mathbf{u}) = \psi_0 - (J-2)\sum_{j=1}^{J}\psi_j(u_j) + \sum_{j=1}^{J-1}\sum_{j'=j+1}^{J}\psi_{jj'}(u_j, u_{j'}), \tag{4.27}$$

where

$$\begin{aligned}
\psi_0 &:= \frac{(J-1)(J-2)}{2}\, R(0,0,\ldots,0) \\
\psi_j(u_j) &:= R(0,\ldots,0,u_j,0,\ldots,0) \\
\psi_{jj'}(u_j,u_{j'}) &:= R(0,\ldots,0,u_j,0,\ldots,0,u_{j'},0,\ldots,0)\,.
\end{aligned} \tag{4.28}$$

By taking conditional and unconditional expectations on both sides of Eq. (4.27), and by using the statistical independence of the biochemical factors, we obtain

$$\begin{aligned}
e_0 &:= \mathrm{E}\big[\widehat{R}(\mathbf{U})\big] \\
&= \psi_0 - (J-2)\sum_{m=1}^{J}\mathrm{E}\big[\psi_m(u_m)\big] \;+\; \sum_{m=1}^{J-1}\sum_{m'=m+1}^{J}\mathrm{E}\big[\psi_{mm'}(U_m,U_{m'})\big] \\
e_j(u_j) &:= \mathrm{E}\big[\widehat{R}(\mathbf{U}) \mid U_j = u_j\big] \\
&= \psi_0 - (J-2)\sum_{m=1}^{J}\mathrm{E}\big[\psi_m(U_m) \mid U_j = u_j\big] \\
&\qquad + \sum_{m=1}^{J-1}\sum_{m'=m+1}^{J}\mathrm{E}\big[\psi_{mm'}(U_m,U_{m'}) \mid U_j = u_j\big]
\end{aligned}$$

$$
\begin{aligned}
e_{jj'}(u_j, u_{j'}) &:= \mathrm{E}\big[R(\mathbf{U}) \mid U_j = u_j, U_{j'} = u_{j'}\big] \\
&= \psi_0 - (J-2)\sum_{m=1}^{J} \mathrm{E}\big[\psi_m(U_m) \mid U_j = u_j, U_{j'} = u_{j'}\big] \\
&\quad + \sum_{m=1}^{J-1}\sum_{m'=m+1}^{J} \mathrm{E}\big[\psi_{mm'}(U_m, U_{m'}) \mid U_j = u_j, U_{j'} = u_{j'}\big], \quad (4.29)
\end{aligned}
$$

where

$$
\mathrm{E}[\psi_m(U_m) \mid U_j = u_j] = \begin{cases} \psi_j(u_j), & \text{if } \ m = j \\ \mathrm{E}[\psi_m(U_m)], & \text{otherwise} \end{cases}
$$

$$
\mathrm{E}[\psi_m(U_m) \mid U_j = u_j, U_{j'} = u_{j'}] = \begin{cases} \psi_j(u_j), & \text{if } \ m = j \\ \psi_{j'}(u_{j'}), & \text{if } \ m = j', \quad \text{for } \ j < j' \\ \mathrm{E}[\psi_m(U_m)], & \text{otherwise} \end{cases}
$$

$$
\mathrm{E}[\psi_{mm'}(U_m, U_{m'}) \mid U_j = u_j] = \begin{cases} \mathrm{E}[\psi_{mj}(U_m, u_j)], & \text{if } \ m < j,\ m' = j \\ \mathrm{E}[\psi_{jm'}(u_j, U_{m'})], & \text{if } \ m = j,\ m' > j \\ \mathrm{E}[\psi_{mm'}(U_m, U_{m'})], & \text{otherwise} \end{cases}
$$

$$
\mathrm{E}[\psi_{mm'}(U_m, U_{m'}) \mid U_j = u_j, U_{j'} = u_{j'}] = \begin{cases} \psi_{jj'}(u_j, u_{j'}), & \text{if } \ m = j,\ m' = j' \\ \mathrm{E}[\psi_{mj}(U_m, u_j)], & \text{if } \ m < j, m' = j \\ \mathrm{E}[\psi_{mj'}(U_m, u_{j'})], & \text{if } \ m \neq j, m' = j' \\ \mathrm{E}[\psi_{jm'}(u_j, U_{m'})], & \text{if } \ m = j, m' \neq j' \\ \mathrm{E}[\psi_{j'm'}(u_{j'}, U_{m'})], & \text{if } \ m = j', m' > j' \\ \mathrm{E}[\psi_{mm'}(U_m, U_{m'})], & \text{otherwise} \end{cases},
$$

for $\ j < j'$. (4.30)

Finally, to compute the conditional variances of the response function $\widehat{R}$, note that

$$\begin{aligned}
&\mathrm{Var}\big[\mathrm{E}\big[\widehat{R}(\mathbf{U}) \mid U_j\big]\big] = \mathrm{Var}\big[e_j(U_j)\big] = \mathrm{E}\big[e_j^2(U_j)\big] - e_0^2, \\
&\mathrm{Var}\big[\mathrm{E}\big[\widehat{R}(\mathbf{U}) \mid U_j, U_{j'}\big]\big] = \mathrm{Var}\big[e_{jj'}(U_j, U_{j'})\big] = \mathrm{E}\big[e_{jj'}^2(U_j, U_{j'})\big] - e_0^2,
\end{aligned} \tag{4.31}$$

since

$$\begin{aligned}
&\mathrm{E}\big[e_j(U_j)\big] = \mathrm{E}\big[\mathrm{E}\big[\widehat{R}(\mathbf{U}) \mid U_j\big]\big] = \mathrm{E}\big[\widehat{R}(\mathbf{U})\big], \\
&\mathrm{E}\big[e_{jj'}(U_j, U_{j'})\big] = \mathrm{E}\big[\mathrm{E}\big[\widehat{R}(\mathbf{U}) \mid U_j, U_{j'}\big]\big] = \mathrm{E}\big[\widehat{R}(\mathbf{U})\big],
\end{aligned} \tag{4.32}$$

by virtue of the fact that $\mathrm{E}\big[\mathrm{E}\big[Y|X\big]\big] = \mathrm{E}\big[Y\big]$.

As a consequence of Eqs. (4.1)–(4.4), and Eq. (4.31), we now obtain the following approximations to the (second-order) SESI's and JESI's:

$$\boxed{\begin{aligned}
\widehat{\sigma}_j^{(2)} &= \frac{\widehat{V}_j}{\widehat{V}}, \qquad \widehat{\eta}_j^{(2)} = \frac{\widehat{T}_j}{\widehat{V}} \\
\widehat{V}_j &= \mathrm{E}\big[e_j^2(U_j)\big] - e_0^2 \\
\widehat{V}_{jj'} &= \mathrm{E}\big[e_{jj'}^2(U_j, U_{j'})\big] - \widehat{V}_j - \widehat{V}_{j'} - e_0^2 \\
\widehat{T}_j &= \sum_{j'=1, j' \neq j}^{J} \widehat{V}_{jj'} \\
\widehat{V} &= \sum_{j=1}^{J} \widehat{V}_j + \sum_{j=1}^{J-1} \sum_{j'=j+1}^{J} \widehat{V}_{jj'}
\end{aligned}} \tag{4.33}$$

with e_0, e_j, and $e_{jj'}$ given by Eq. (4.29). Note that evaluation of the expectations of these quantities requires only one- and two-dimensional integrations. This can be done by a standard Gauss-Hermite integration procedure, as we explain in Subsection 4.5.3. We respectively refer to $\widehat{\sigma}_j^{(2)}$ and $\widehat{\eta}_j^{(2)}$, given by Eq. (4.33), as the (second-order) SESI's and JESI's obtained by *Gauss-Hermite Integration* (GHI) of the FD-HDMR.

4.4.4 OHA of ANOVA-HDMR

Eq. (4.11) and the fact that the integrals of basis functions p over a *single* variable are equal to zero imply

$$\begin{aligned}\int_0^1\int_0^1\int_0^1 P^2(s_1,s_2,s_3)ds_1ds_2ds_3 = p_0^2 &+ \int_0^1 p_1^2(s_1)ds_1 + \int_0^1 p_2^2(s_2)ds_2 \\ &+ \int_0^1 p_3^2(s_3)ds_3 \\ &+ \int_0^1\int_0^1 p_{12}^2(s_1,s_2)ds_1ds_2 \\ &+ \int_0^1\int_0^1 p_{13}^2(s_1,s_3)ds_1ds_3 \\ &+ \int_0^1\int_0^1 p_{23}^2(s_2,s_3)ds_2ds_3 \\ &+ \int_0^1\int_0^1\int_0^1 p_{123}^2(s_1,s_2,s_3)ds_1ds_2ds_3\,.\end{aligned} \tag{4.34}$$

If we assume that the biochemical factors of interest are statistically independent random variables U_1, U_2, and U_3, with cumulative distribution functions $g_1(u_1)$, $g_2(u_2)$, and $g_3(u_3)$, respectively, then Eq. (4.34), together with Eqs. (4.10), (4.12), and (4.14), implies that

$$V = V_1 + V_2 + V_3 + V_{12} + V_{13} + V_{23} + V_{123}, \tag{4.35}$$

where

$$\begin{aligned}V &:= \mathrm{Var}\big[R(U_1,U_2,U_3)\big] \\ V_1 &:= \mathrm{Var}\big[\mathrm{E}\big[R(U_1,U_2,U_3)\mid U_1\big]\big] = \int_{-\infty}^{\infty}\rho_1^2(u_1)g_1'(u_1)du_1 \\ V_2 &:= \mathrm{Var}\big[\mathrm{E}\big[R(U_1,U_2,U_3)\mid U_2\big]\big] = \int_{-\infty}^{\infty}\rho_2^2(u_2)g_2'(u_2)du_2 \\ V_3 &:= \mathrm{Var}\big[\mathrm{E}\big[R(U_1,U_2,U_3)\mid U_3\big]\big] = \int_{-\infty}^{\infty}\rho_3^2(u_3)g_3'(u_3)du_3 \\ V_{12} &:= \mathrm{Var}\big[\mathrm{E}\big[R(U_1,U_2,U_3)\mid U_1,U_2\big]\big] - V_1 - V_2 \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\rho_{12}^2(u_1,u_2)g_1'(u_1)g_2'(u_2)du_1du_2 \geq 0\end{aligned}$$

$$
\begin{aligned}
V_{13} &:= \mathrm{Var}\big[\mathrm{E}\big[R(U_1, U_2, U_3) \mid U_1, U_3\big]\big] - V_1 - V_3 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \rho_{13}^2(u_1, u_3) g_1'(u_1) g_3'(u_3) du_1 du_3 \geq 0 \\
V_{23} &:= \mathrm{Var}\big[\mathrm{E}\big[R(U_1, U_2, U_3) \mid U_2, U_3\big]\big] - V_2 - V_3 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \rho_{23}^2(u_2, u_3) g_2'(u_2) g_3'(u_3) du_2 du_3 \geq 0 \\
V_{123} &:= V - V_{12} - V_{13} - V_{23} - V_1 - V_2 - V_3 \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \rho_{123}^2(u_1, u_2, u_3) g_1'(u_1) g_2'(u_2) g_3'(u_3) du_1 du_2 du_3 \geq 0, \quad (4.36)
\end{aligned}
$$

since $g_1'(u_1)$, $g_2'(u_2)$, and $g_3'(u_3)$ are the probability density functions of U_1, U_2, and U_3, respectively. The variance decomposition scheme given by Eqs. (3.1) and (3.2) is a general version of the decomposition given by Eqs. (4.35) and (4.36) for the case of J biochemical factors. This decomposition is closely related to *analysis of variance* (ANOVA) techniques in statistics [26, 80, 81]. For this reason, the representation of the response function R by Eqs. (4.13) and (4.14) is referred to in the literature as ANOVA-HDMR.

Note that Eq. (4.35) can be shown in a trivial manner by adding all V's in Eq. (4.36). However, by using Eqs. (4.10), (4.12), (4.14), and (4.36), we can show that, when U_1, U_2, and U_3 are statistically independent, then $V_{12}, V_{13}, V_{23}, V_{123} \geq 0$, which is a crucial property for appropriately defining the variance-based sensitivity indices we consider in this dissertation. Moreover, we can show that these quantities can be directly evaluated from the basis functions of the ANOVA-HDMR of the response function $R(\mathbf{u})$ by means of Eq. (4.36). As a consequence, we can use ANOVA-HDMR to develop an efficient approximation technique for the sensitivity indices $\sigma_j^{(2)}$ and $\eta_j^{(2)}$. We can do this by

sufficiently approximating the response $R(u_1, u_2, u_3)$ by a function

$$\widehat{R}(u_1, u_2, u_3) = \widehat{\rho}_0 + \widehat{\rho}_1(u_1) + \widehat{\rho}_2(u_2) + \widehat{\rho}_3(u_3) + \widehat{\rho}_{12}(u_1, u_2) + \widehat{\rho}_{13}(u_1, u_3) + \widehat{\rho}_{23}(u_2, u_3), \tag{4.37}$$

where the approximating basis functions $\widehat{\rho}$ must be appropriately chosen so that they satisfy the necessary orthogonality conditions, given by Eq. (4.15), and allow efficient evaluation of the integrals in Eq. (4.36).

There are several potential choices for the approximating basis functions $\widehat{\rho}$, such as polynomials, exponentials, splines, etc. However, for the case of statistically independent zero-mean Gaussian biochemical factors, the simplest choice is based on the following first- and second-order Hermite polynomials:

$$H_1(x) = x \qquad \text{and} \qquad H_2(x) = \frac{x^2 - 1}{\sqrt{2}}\,. \tag{4.38}$$

Note that these polynomials are orthonormal over the standard Gaussian distribution, satisfying

$$\begin{aligned}
&\int_{-\infty}^{\infty} H_1(x)\,\frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}dx = \int_{-\infty}^{\infty} H_2(x)\,\frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}dx = 0\\
&\int_{-\infty}^{\infty} H_1^2(x)\,\frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}dx = \int_{-\infty}^{\infty} H_2^2(x)\,\frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}dx = 1\\
&\int_{-\infty}^{\infty} H_1(x)H_2(x)\,\frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}dx = 0.
\end{aligned} \tag{4.39}$$

In this case, we set

$$\begin{aligned}
\widehat{\rho}_j(u_j) \;&=\; \alpha_{j,1}\frac{u_j}{\lambda_j} + \frac{\alpha_{j,2}}{\sqrt{2}}\left(\frac{u_j^2}{\lambda_j^2} - 1\right)\\
\widehat{\rho}_{jj'}(u_j, u_{j'}) \;&=\; \alpha_{jj',1}\frac{u_j u_{j'}}{\lambda_j \lambda_{j'}} + \frac{\alpha_{jj',2}}{\sqrt{2}}\left(\frac{u_j^2}{\lambda_j^2} - 1\right)\frac{u_{j'}}{\lambda_{j'}} + \frac{\alpha_{jj',3}}{\sqrt{2}}\frac{u_j}{\lambda_j}\left(\frac{u_{j'}^2}{\lambda_{j'}^2} - 1\right)\\
&\qquad + \frac{\alpha_{jj',4}}{2}\left(\frac{u_j^2}{\lambda_j^2} - 1\right)\left(\frac{u_{j'}^2}{\lambda_{j'}^2} - 1\right).
\end{aligned} \tag{4.40}$$

Note that, since the biochemical factors U_j are statistically independent zero-mean Gaussian random variables with standard deviations given by λ_j, these approximations satisfy the necessary orthogonality conditions given by Eq. (4.15).

By using Eq. (4.36) and the orthonormality of the Hermite polynomials H_1 and H_2, given by Eq. (4.39), we can obtain the following approximations to the (second-order) SESI's and JESI's (expressed for the general case of J biochemical factors):

$$\boxed{\begin{gathered} \widehat{\sigma}_j^{(2)} = \frac{\widehat{V}_j}{\widehat{V}}, \qquad \widehat{\eta}_j^{(2)} = \frac{\widehat{T}_j}{\widehat{V}} \\ \widehat{V}_j = \alpha_{j,1}^2 + \alpha_{j,2}^2 \\ \widehat{V}_{jj'} = \alpha_{jj',1}^2 + \alpha_{jj',2}^2 + \alpha_{jj',3}^2 + \alpha_{jj',4}^2 \\ \widehat{T}_j = \sum_{j'=1, j' \neq j}^{J} \widehat{V}_{jj'} \\ \widehat{V} = \sum_{j=1}^{J} \widehat{V}_j + \sum_{j=1}^{J-1} \sum_{j'=j+1}^{J} \widehat{V}_{jj'} \end{gathered}} \tag{4.41}$$

We respectively refer to $\widehat{\sigma}_j^{(2)}$ and $\widehat{\eta}_j^{(2)}$, given by Eq. (4.41), as the (second-order) SESI's and JESI's obtained by *Orthonormal Hermite Approximation* (OHA) of the ANOVA-HDMR.

4.5 Numerical Implementation

We now discuss the numerical implementation of the approximation techniques we presented in Section 4.4. Some techniques can be implemented in a straightforward manner, while others require more involved implementation steps.

4.5.1 Derivative Approximation

Approximation of the SESI's and JESI's by means of Eq. (4.20) requires evaluation of the first- and second-order partial derivatives of the response function $R(\mathbf{u})$ at $\mathbf{u} = \mathbf{0}$, given by

$$d_j = \frac{\partial R(\mathbf{0})}{\partial u_j} \qquad \text{and} \qquad d_{jj'} = \frac{\partial^2 R(\mathbf{0})}{\partial u_j \partial u_{j'}} \,. \tag{4.42}$$

Unfortunately, accurate evaluation of these derivatives is not an easy task [82]. We may express them in terms of concentration sensitivities and analytically derive a system of differential equations that govern the dynamic evolution of such sensitivities. Then, evaluation of the response derivatives will require simultaneous integration of the sensitivity equations together with the differential equations governing the underlying molecular concentration dynamics. Most often, this step cannot be implemented in a reasonable time due to stiffness of the resulting differential equations [4]. As a consequence, the derivatives are usually approximated by finite-differences. However, the resulting approximations must be carefully used, since it is difficult to theoretically predict, control, and numerically evaluate the accuracy of finite-difference approximations of derivatives [4].

In this dissertation, we use symmetric finite-difference approximations of the derivatives. A symmetric finite-difference approximation of the first-order partial derivative d_j of $R(\mathbf{u})$ with respect to u_j at $\mathbf{0}$, leads to

$$d_j \simeq \frac{R(\Delta \mathbf{e}_j) - R(-\Delta \mathbf{e}_j)}{2\Delta} \,, \tag{4.43}$$

for a sufficiently small differential step size $\Delta > 0$, where $\mathbf{e}_j$ denotes a J-dimensional vector with its j^{th} element being equal to one and the remaining elements being zero.

By applying the previous equation twice, we obtain the following finite-difference approximation for the second-order partial derivative $d_{jj'}$ of $R(\mathbf{u})$ with respect to u_j and $u_{j'}$ at $\mathbf{0}$:

$$d_{jj'} \simeq \frac{R(\Delta\mathbf{e}_j + \Delta\mathbf{e}_{j'}) - R(-\Delta\mathbf{e}_j + \Delta\mathbf{e}_{j'}) - R(\Delta\mathbf{e}_j - \Delta\mathbf{e}_{j'}) + R(-\Delta\mathbf{e}_j - \Delta\mathbf{e}_{j'})}{4\Delta^2}. \quad (4.44)$$

To compute these approximations, we need $2J(J+1)+1$ system integrations, which is quadratic in terms of the number J of the underlying biochemical factors and is much smaller than the number $2L(J+1)$ of system integrations required by Monte Carlo (MC) estimation, since $J \ll L$.

4.5.2 Polynomial Approximation of FD-HDMR

The approximation of the SESI's and JESI's by means of Eq. (4.24) requires knowledge of the values of the α parameters associated with the polynomial approximation of the basis functions r, given by Eq. (4.23). This can be done by polynomial regression [83], as we explain next.

Our problem here is to estimate the parameters α, so that

$$r_j(u_j) = \widehat{r}_j(u_j) + \epsilon_j = \alpha_{j,1}u_j + \alpha_{j,2}u_j^2 + \epsilon_j, \quad (4.45)$$

and

$$\begin{aligned} r_{jj'}(u_j, u_{j'}) &= \widehat{r}_{jj'}(u_j, u_{j'}) + \epsilon_{jj'} \\ &= \alpha_{jj',1}u_j u_{j'} + \alpha_{jj',2}u_j^2 u_{j'} + \alpha_{jj',3}u_j u_{j'}^2 + \alpha_{jj',4}u_j^2 u_{j'}^2 + \epsilon_{jj'}, \end{aligned} \quad (4.46)$$

for every j, j', where the ϵ's are zero-mean random variables that model the errors of approximating the basis functions r by $\widehat{r}$. We can now use Eq. (4.9) to evaluate the

basis functions r at a set $\{u_j(q), q \in S, j \in J\}$ of pre-specified factor values around zero. Then, the least-square error estimates $\widehat{\alpha}_{j,1}, \widehat{\alpha}_{j,2}$ of the parameters $\alpha_{j,1}, \alpha_{j,2}$ associated with the basis function $r_j(u_j)$ are given by [83]:

$$\widehat{\boldsymbol{\alpha}}_j = (\mathbb{U}_j^T \, \mathbb{U}_j)^{-1} \mathbb{U}_j^T \, \mathbf{r}_j \, , \tag{4.47}$$

where

$$\widehat{\boldsymbol{\alpha}}_j := \begin{bmatrix} \widehat{\alpha}_{j,1} \\ \\ \widehat{\alpha}_{j,2} \end{bmatrix}_{2\times 1} , \qquad \mathbf{r}_j := \begin{bmatrix} r_j(u_j(1)) \\ \\ r_j(u_j(2)) \\ \\ \vdots \\ \\ r_j(u_j(S)) \end{bmatrix}_{S\times 1} , \qquad \mathbb{U}_j := \begin{bmatrix} u_j(1) & u_j^2(1) \\ \\ u_j(2) & u_j^2(2) \\ \\ \vdots & \vdots \\ \\ u_j(S) & u_j^2(S) \end{bmatrix}_{S\times 2} , \tag{4.48}$$

provided that the matrix $\mathbb{U}_j^T \mathbb{U}_j$ is invertible (which is always true if no column of the $\mathbb{U}_j$ matrix is a linear combination of the other columns). On the other hand, the least-square error estimates $\widehat{\alpha}_{jj',1}$, $\widehat{\alpha}_{jj',2}$, $\widehat{\alpha}_{jj',3}$, $\widehat{\alpha}_{jj',4}$ of the parameters $\alpha_{jj',1}$, $\alpha_{jj',2}$, $\alpha_{jj',3}, \alpha_{jj',4}$ associated with the basis function $\widehat{r}_{jj'}(u_j, u_{j'})$ are given by [83]:

$$\widehat{\boldsymbol{\alpha}}_{jj'} = (\mathbb{U}_{jj'}^T \, \mathbb{U}_{jj'})^{-1} \mathbb{U}_{jj'}^T \, \mathbf{r}_{jj'} \, , \tag{4.49}$$

where

$$\widehat{\boldsymbol{\alpha}}_{jj'} := \begin{bmatrix} \widehat{\alpha}_{jj',1} \\ \\ \widehat{\alpha}_{jj',2} \\ \\ \widehat{\alpha}_{jj',3} \\ \\ \widehat{\alpha}_{jj',4} \end{bmatrix}_{4\times 1}$$

$$\mathbf{r}_{jj'} := \begin{bmatrix} r_{jj'}(u_j(1), u_{j'}(1)) \\ \vdots \\ r_{jj'}(u_j(1), u_{j'}(S)) \\ r_{jj'}(u_j(2), u_{j'}(1)) \\ \vdots \\ r_{jj'}(u_j(2), u_{j'}(S)) \\ \vdots \\ r_{jj'}(u_j(S), u_{j'}(1)) \\ \vdots \\ r_{jj'}(u_j(S), u_{j'}(S)) \end{bmatrix}_{S^2 \times 1}$$

$$\mathbb{U}_{jj'} := \begin{bmatrix} u_j(1)u_{j'}(1) & u_j^2(1)u_{j'}(1) & u_j(1)u_{j'}^2(1) & u_j^2(1)u_{j'}^2(1) \\ \vdots & \vdots & \vdots & \vdots \\ u_j(1)u_{j'}(S) & u_j^2(1)u_{j'}(S) & u_j(1)u_{j'}^2(S) & u_j^2(1)u_{j'}^2(S) \\ u_j(2)u_{j'}(1) & u_j^2(2)u_{j'}(1) & u_j(2)u_{j'}^2(1) & u_j^2(2)u_{j'}^2(1) \\ \vdots & \vdots & \vdots & \vdots \\ u_j(2)u_{j'}(S) & u_j^2(2)u_{j'}(S) & u_j(2)u_{j'}^2(S) & u_j^2(2)u_{j'}^2(S) \\ \vdots & \vdots & \vdots & \vdots \\ u_j(S)u_{j'}(1) & u_j^2(S)u_{j'}(1) & u_j(S)u_{j'}^2(1) & u_j^2(S)u_{j'}^2(1) \\ \vdots & \vdots & \vdots & \vdots \\ u_j(S)u_{j'}(S) & u_j^2(S)u_{j'}(S) & u_j(S)u_{j'}^2(S) & u_j^2(S)u_{j'}^2(S) \end{bmatrix}_{S^2 \times 4}, \tag{4.50}$$

provided that the matrix $\mathbb{U}_{jj'}^T \mathbb{U}_{jj'}$ is invertible. Note that calculation of $\widehat{\boldsymbol{\alpha}}$ requires $J(J-1)S^2/2 + JS + 1$ system integrations, which is quadratic both in terms of the number J of biochemical factors and the number S of the samples per factor used in the regression. Note that $J(J-1)S^2/2 + JS + 1 \simeq 2J^2(S/2)^2$, for sufficiently large J. This number is much smaller than the number $2L(J+1) \simeq 2LJ$ of system integrations required by MC, since $L \gg J(S/2)^2$, but larger than the number $2J(J+1)+1 \simeq 2J^2$ of system integrations required by Derivative Approximation (DA), since $S > 2$.

4.5.3 Gauss-Hermite Integration of FD-HDMR

It is clear from Eqs. (4.29) and (4.30) that evaluation of the SESI's and JESI's by Eq. (4.33) requires the calculation of $\mathrm{E}\big[\psi_m(U_m)\big]$, $\mathrm{E}\big[\psi_{jm'}(U_j, U_m)\big]$, $\mathrm{E}\big[\psi_{mj}(U_m, U_j)\big]$, $\mathrm{E}\big[\psi_{mm'}(U_m, U_{m'})\big]$, $\mathrm{E}\big[e_j^2(u_j)\big]$, and $\mathrm{E}\big[e_{jj'}^2(u_j, u_{j'})\big]$ with respect to Gaussian distributions. We can evaluate these expectations by using Gauss-Hermite integration [84], as we explain next.

Let us consider the one-dimensional expectation:

$$E_1 = \mathrm{E}\big[\psi_1(U_1)\big] = \frac{1}{\lambda\sqrt{2\pi}} \int_{-\infty}^{\infty} \psi_1(U_1) e^{-u_1^2/2\lambda^2} du_1. \tag{4.51}$$

If we set $u_1 = \sqrt{2}\lambda v_1$, then

$$E_1 = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \psi_1(\sqrt{2}\lambda v_1) e^{-v_1^2} dv_1. \tag{4.52}$$

In this form, we can use the Gauss-Hermite integration procedure to approximate E_1 by

$$\widehat{E}_1 = \frac{1}{\sqrt{\pi}} \sum_{q=1}^{Q} \omega_q \psi_1(\sqrt{2}\lambda a_q), \tag{4.53}$$

where Q is the order of the approximation, a_q and ω_q are appropriately chosen abscissas and weights, respectively [84].

Likewise, by setting $u_1 = \sqrt{2}\lambda_1 v_1$ and $u_2 = \sqrt{2}\lambda_2 v_2$, we can write the two-dimensional expectation

$$E_2 = \mathrm{E}\big[\psi_2(U_1, U_2)\big] = \frac{1}{2\pi\lambda_1\lambda_2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\psi_2(u_1, u_2)e^{-u_1^2/2\lambda_1^2}e^{-u_2^2/2\lambda_2^2}du_1du_2 \tag{4.54}$$

in the form

$$E_2 = \frac{1}{\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\psi_2(\sqrt{2}\lambda_1 v_1, \sqrt{2}\lambda_2 v_2)e^{-v_1^2}e^{-v_2^2}dv_1dv_2. \tag{4.55}$$

A two-step (first for v_1 and then for v_2) application of one-dimensional Gauss-Hermite integration results in the following approximation of E_2:

$$\widehat{E}_2 = \frac{1}{\pi}\sum_{q_1=1}^{Q}\sum_{q_2=1}^{Q}\omega_{q_1}\omega_{q_2}\psi_2(\sqrt{2}\lambda_1 a_{q_1}, \sqrt{2}\lambda_2 a_{q_2}). \tag{4.56}$$

It turns out that calculation of the expectations required by Eq. (4.33) involves $J(J-1)Q^2/2+ JQ + 1$ system integrations, when Q is even, or $J(J-1)(Q-1)^2/2 + J(Q-1) + 1$ system integrations, when Q is odd, which is quadratic both in terms of the number J of biochemical factors and the order Q of Gauss-Hermite Integration (GHI). Note that, if the number S of the samples per factor used in the regression associated with Polynomial Approximation (PA) is even, and $Q = S$ or $Q = S + 1$, then GHI requires the same number of system integrations as PA.

4.5.4 OHA of ANOVA-HDMR

Approximating the sensitivity indices $\sigma_j^{(2)}$ and $\eta_j^{(2)}$ by Eq. (4.41) requires evaluation of the parameters α so that the functions $\widehat{\rho}$, given by Eq. (4.40), result in a sufficiently

good approximation of the response function R by $\widehat{R}$, given by Eq. (4.37). Our problem here is to estimate the parameters α, so that

$$\rho_j(u_j) = \widehat{\rho}_j(u_j) + \epsilon_j = \alpha_{j,1}\frac{u_j}{\lambda_j} + \frac{\alpha_{j,2}}{\sqrt{2}}\left(\frac{u_j^2}{\lambda_j^2} - 1\right) + \epsilon_j, \tag{4.57}$$

and

$$\begin{aligned}\rho_{jj'}(u_j, u_{j'}) &= \widehat{\rho}_{jj'}(u_j, u_{j'}) + \epsilon_{jj'} \\ &= \alpha_{jj',1}\frac{u_j u_{j'}}{\lambda_j \lambda_{j'}} + \frac{\alpha_{jj',2}}{\sqrt{2}}\left(\frac{u_j^2}{\lambda_j^2} - 1\right)\frac{u_{j'}}{\lambda_{j'}} + \frac{\alpha_{jj',3}}{\sqrt{2}}\frac{u_j}{\lambda_j}\left(\frac{u_{j'}^2}{\lambda_{j'}^2} - 1\right) \\ &\quad + \frac{\alpha_{jj',4}}{2}\left(\frac{u_j^2}{\lambda_j^2} - 1\right)\left(\frac{u_{j'}^2}{\lambda_{j'}^2} - 1\right) + \epsilon_{jj'}\,,\end{aligned} \tag{4.58}$$

for every j, j', where the ϵ's are zero-mean random variables that model the errors in approximating the basis functions ρ by $\widehat{\rho}$. From Eq. (4.57), note that

$$\begin{aligned}\alpha_{j,1} &= \int_{-\infty}^{\infty} \frac{u_j}{\lambda_j}\rho_j(u_j)G_j(u_j)du_j, \\ \alpha_{j,2} &= \frac{1}{\sqrt{2}}\int_{-\infty}^{\infty}\left(\frac{u_j^2}{\lambda_j^2} - 1\right)\rho_j(u_j)G_j(u_j)du_j,\end{aligned} \tag{4.59}$$

where $G_j(u_j)$ is the Gaussian probability density function

$$G_j(u_j) = \frac{1}{\sqrt{2\pi}\lambda_j}\, e^{-u_j^2/2\lambda_j^2}. \tag{4.60}$$

This is a consequence of the zero-mean Gaussianity of the biochemical factors and the orthonormality of the Hermite polynomials over the Gaussian distribution. Likewise, and from Eq. (4.58), we have that

$$\alpha_{jj',1} = \int_{-\infty}^{\infty} \frac{u_j}{\lambda_j}\frac{u_{j'}}{\lambda_{j'}}\,\rho_{jj'}(u_j, u_{j'})G_j(u_j)G_{j'}(u_{j'})du_j du_{j'}$$

$$\alpha_{jj',2} = \frac{1}{\sqrt{2}}\int_{-\infty}^{\infty}\left(\frac{u_j^2}{\lambda_j^2} - 1\right)\frac{u_{j'}}{\lambda_{j'}}\,\rho_{jj'}(u_j, u_{j'})G_j(u_j)G_{j'}(u_{j'})du_j du_{j'}$$

$$\alpha_{jj',3} = \frac{1}{\sqrt{2}} \int_{-\infty}^{\infty} \frac{u_j}{\lambda_j} \left(\frac{u_{j'}^2}{\lambda_{j'}^2} - 1 \right) \rho_{jj'}(u_j, u_{j'}) G_j(u_j) G_{j'}(u_{j'}) du_j du_{j'}$$

$$\alpha_{jj',4} = \frac{1}{2} \int_{-\infty}^{\infty} \left(\frac{u_j^2}{\lambda_j^2} - 1 \right) \left(\frac{u_{j'}^2}{\lambda_{j'}^2} - 1 \right) \rho_{jj'}(u_j, u_{j'}) G_j(u_j) G_{j'}(u_{j'}) du_j du_{j'}. \quad (4.61)$$

Finally,

$$\begin{aligned}
\alpha_{j,1} &= \mathrm{E}\left[\frac{U_j}{\lambda_j} R(\mathbf{U}) \right] \\
\alpha_{j,2} &= \frac{1}{\sqrt{2}} \mathrm{E}\left[\left(\frac{U_j^2}{\lambda_j^2} - 1 \right) R(\mathbf{U}) \right] \\
\alpha_{jj',1} &= \mathrm{E}\left[\frac{U_j}{\lambda_j} \frac{U_{j'}}{\lambda_{j'}} R(\mathbf{U}) \right] \\
\alpha_{jj',2} &= \frac{1}{\sqrt{2}} \mathrm{E}\left[\left(\frac{U_j^2}{\lambda_j^2} - 1 \right) \frac{U_{j'}}{\lambda_{j'}} R(\mathbf{U}) \right] \\
\alpha_{jj',3} &= \frac{1}{\sqrt{2}} \mathrm{E}\left[\frac{U_j}{\lambda_j} \left(\frac{U_{j'}^2}{\lambda_{j'}^2} - 1 \right) R(\mathbf{U}) \right] \\
\alpha_{jj',4} &= \frac{1}{2} \mathrm{E}\left[\left(\frac{U_j^2}{\lambda_j^2} - 1 \right) \left(\frac{U_{j'}^2}{\lambda_{j'}^2} - 1 \right) R(\mathbf{U}) \right], \quad (4.62)
\end{aligned}$$

by virtue of Eqs. (4.14), (4.59), and (4.61).

As a consequence of the previous analysis, to determine the parameters α, we need to evaluate the expectations in Eq. (4.62). We can do this by Monte Carlo estimation based on a Latin hypercube sampling strategy, which leads to a more efficient implementation than standard Monte Carlo sampling [66, 67]. In particular, we can generate L Latin hypercube Gaussian samples $\mathbf{u}^{(l)} = \{u_1^{(l)}, u_2^{(l)}, \ldots, u_J^{(l)}\}$, $l = 1, 2, \ldots, L$, evaluate the responses $R(\mathbf{u}^{(l)})$, for $l = 1, 2, \ldots, L$, and set

$$\alpha_{j,1} \simeq \widehat{\alpha}_{j,1} := \frac{1}{L} \sum_{l=1}^{L} \frac{u_j^{(l)}}{\lambda_j} R(\mathbf{u}^{(l)})$$

$$\alpha_{j,2} \simeq \widehat{\alpha}_{j,2} := \frac{1}{\sqrt{2}} \frac{1}{L} \sum_{l=1}^{L} \left(\frac{\left[u_j^{(l)} \right]^2}{\lambda_j^2} - 1 \right) R(\mathbf{u}^{(l)})$$

$$\alpha_{jj',1} \simeq \widehat{\alpha}_{jj',1} := \frac{1}{L}\sum_{l=1}^{L} \frac{u_j^{(l)}}{\lambda_j}\frac{u_{j'}^{(l)}}{\lambda_{j'}} R(\mathbf{u}^{(l)})$$

$$\alpha_{jj',2} \simeq \widehat{\alpha}_{jj',2} := \frac{1}{\sqrt{2}}\,\frac{1}{L}\sum_{l=1}^{L} \left(\frac{\left[u_j^{(l)}\right]^2}{\lambda_j^2} - 1 \right) \frac{u_{j'}^{(l)}}{\lambda_{j'}} R(\mathbf{u}^{(l)})$$

$$\alpha_{jj',3} \simeq \widehat{\alpha}_{jj',3} := \frac{1}{\sqrt{2}}\,\frac{1}{L}\sum_{l=1}^{L} \frac{u_j^{(l)}}{\lambda_j} \left(\frac{\left[u_{j'}^{(l)}\right]^2}{\lambda_{j'}^2} - 1 \right) R(\mathbf{u}^{(l)})$$

$$\alpha_{jj',4} \simeq \widehat{\alpha}_{jj',4} := \frac{1}{2}\,\frac{1}{L}\sum_{l=1}^{L} \left(\frac{\left[u_j^{(l)}\right]^2}{\lambda_j^2} - 1 \right) \left(\frac{\left[u_{j'}^{(l)}\right]^2}{\lambda_{j'}^2} - 1 \right) R(\mathbf{u}^{(l)}). \qquad (4.63)$$

Clearly, implementation of Eq. (4.63) requires L system integrations.

The problem with Monte Carlo estimation is that, most often, it requires a large number of system integrations to produce sufficiently accurate estimates for the α parameters. As a consequence, it is a computationally inefficient method for estimating α. An alternative approach is to use the previous L samples $\{\mathbf{u}^{(l)}, l = 1, 2, \ldots, L\}$ and estimate the α parameters by polynomial regression, as we did in Subsection 4.5.2. We discuss this approach in the following.

As a consequence of Eqs. (4.37) and (4.40), the polynomial regression problem amounts to estimating $\widehat{\rho}_0$ and the parameters α, so that

$$\begin{aligned} R(\mathbf{u}) \;&=\; \widehat{R}(\mathbf{u}) + \epsilon \\ &=\; \widehat{\rho}_0 + \sum_{j=1}^{J} \alpha_{j,1}\frac{u_j}{\lambda_j} + \frac{\alpha_{j,2}}{\sqrt{2}}\left(\frac{u_j^2}{\lambda_j^2} - 1\right) \\ &\quad + \sum_{j=1}^{J-1}\sum_{j'=j+1}^{J} \alpha_{jj',1}\frac{u_j u_{j'}}{\lambda_j \lambda_{j'}} + \frac{\alpha_{jj',2}}{\sqrt{2}}\left(\frac{u_j^2}{\lambda_j^2} - 1\right)\frac{u_{j'}}{\lambda_{j'}} \\ &\quad + \sum_{j=1}^{J-1}\sum_{j'=j+1}^{J} \frac{\alpha_{jj',3}}{\sqrt{2}}\frac{u_j}{\lambda_j}\left(\frac{u_{j'}^2}{\lambda_{j'}^2} - 1\right) + \frac{\alpha_{jj',4}}{2}\left(\frac{u_j^2}{\lambda_j^2} - 1\right)\left(\frac{u_{j'}^2}{\lambda_{j'}^2} - 1\right) + \epsilon\,, \end{aligned} \qquad (4.64)$$

where ϵ is a zero-mean random variable that models the errors of approximating the

response function R by $\widehat{R}$. In this case, the least-square error estimate $\widehat{\boldsymbol{\alpha}}$ of the parameters α are given by

$$\widehat{\boldsymbol{\alpha}} = (\mathbb{U}^T\mathbb{U})^{-1}\mathbb{U}^T\rho\,, \tag{4.65}$$

where

$$\widehat{\boldsymbol{\alpha}} := \begin{bmatrix} \widehat{\rho}_0 \\ \widehat{\alpha}_{1,1} \\ \vdots \\ \widehat{\alpha}_{1,2} \\ \vdots \\ \widehat{\alpha}_{(J-1)J,4} \end{bmatrix}_{(2J^2+1)\times 1} \qquad \rho := \begin{bmatrix} R(\mathbf{u}^{(1)}) \\ R(\mathbf{u}^{(2)}) \\ \vdots \\ R(\mathbf{u}^{(L)}) \end{bmatrix}_{L\times 1}$$

$$\mathbb{U} := \begin{bmatrix} 1 & \frac{u_1^{(1)}}{\lambda_1} & \cdots & \frac{\left[(u_1^{(1)}/\lambda_1)^2-1\right]}{\sqrt{2}} & \cdots & \frac{\left[(u_{J-1}^{(1)}/\lambda_{J-1})^2-1\right]\left[(u_J^{(1)}/\lambda_J)^2-1\right]}{2} \\ 1 & \frac{u_1^{(2)}}{\lambda_1} & \cdots & \frac{\left[(u_1^{(2)}/\lambda_1)^2-1\right]}{\sqrt{2}} & \cdots & \frac{\left[(u_{J-1}^{(2)}/\lambda_{J-1})^2-1\right]\left[(u_J^{(2)}/\lambda_J)^2-1\right]}{2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \frac{u_1^{(L)}}{\lambda_1} & \cdots & \frac{\left[(u_1^{(L)}/\lambda_1)^2-1\right]}{\sqrt{2}} & \cdots & \frac{\left[(w_{J-1}^{(L)}/\lambda_{J-1})^2-1\right]\left[(u_J^{(L)}/\lambda_J)^2-1\right]}{2} \end{bmatrix}_{L\times(2J^2+1)}, \tag{4.66}$$

provided that the matrix $\mathbb{U}^T\mathbb{U}$ is invertible. Note that calculation of $\widehat{\boldsymbol{\alpha}}$ requires the same number L of system integrations as Monte Carlo estimation by Eq. (4.63).

It is not difficult to see from Eq. (4.65) that, if $\widehat{\boldsymbol{\alpha}}_{\rm MC}$ is the Monte Carlo estimate of $\widehat{\rho}_0$ and of the parameters α, given by [recall Eq. (4.14)]

$$\widehat{\rho}_0 = \rho_0 = \mathrm{E}\big[R(\mathbf{U})\big] \simeq \frac{1}{L}\sum_{l=1}^{L} R(\mathbf{u}^{(l)}), \tag{4.67}$$

and Eq. (4.63), respectively, then

$$\widehat{\boldsymbol{\alpha}}_{\rm MC} = \frac{1}{L}\,\mathbb{U}^T\rho = \frac{1}{L}\,\mathbb{U}^T\mathbb{U}\,\widehat{\boldsymbol{\alpha}}\,. \tag{4.68}$$

Moreover,

$$\lim_{L\to\infty} \frac{1}{L}\mathbb{U}^T\mathbb{U} = \mathbb{I}, \tag{4.69}$$

where $\mathbb{I}$ is the identity matrix, by virtue of the biorthonormality conditions given by Eq. (4.39) and the fact that the Monte Carlo estimate $\left[\sum_{l=1}^{L} f(x^{(l)})\right]/L$ converges to the integral $\int_{-\infty}^{\infty} f(x)\pi(x)dx$, as $L \to \infty$, provided that $x^{(l)}$, $l = 1, 2, \ldots, L$, are samples independently drawn from the probability density function $\pi(x)$. As a consequence, the Monte Carlo estimate $\widehat{\boldsymbol{\alpha}}_{\mathrm{MC}}$ and the regression estimate $\widehat{\boldsymbol{\alpha}}$ are identical in the limit as the number of Monte Carlo samples grows to infinity. Since α is obtained by minimizing the least-square error between R and $\widehat{R}$, we expect that the regression estimate α of the parameters α will be more preferable than the Monte Carlo estimate $\widehat{\boldsymbol{\alpha}}_{\mathrm{MC}}$, in the sense that, for a relatively small number of Monte Carlo samples, α may produce a better fit $\widehat{R}$ of the response function R than the one produced by $\widehat{\boldsymbol{\alpha}}_{\mathrm{MC}}$. Finally, note from Eq. (4.69) that, for a sufficiently large number of Monte Carlo samples, $\mathbb{U}^T\mathbb{U}$ is approximately equal to the identity matrix multiplied by L, which effectively reduces the risk of singularity when evaluating the inverse matrix $(\mathbb{U}^T\mathbb{U})^{-1}$ in Eq. (4.65). Therefore, if $\mathbb{U}^T\mathbb{U}$ turns out to be singular for a chosen value of L, the user needs to increase L until a nonsingular matrix $\mathbb{U}^T\mathbb{U}$ is obtained.

Orthonormal Hermite Approximation (OHA) requires L system integrations, where L is the number of regression points obtained by Latin hypercube sampling.[7] This number is smaller than the number $2L(J+1)$ of system integrations used in MC by a factor of $2(J+1)$, but it could be larger than the number of system integrations required by DA, PA, or GHI.

[7] In this chapter, we take the number of regression points used in OHA to be the same as the number of Latin hypercube samples employed by MC, although these two numbers can be different in general.

4.6 Numerical Results

We now employ the techniques discussed in this chapter to estimate the variance-based sensitivity indices $\sigma_{j,2}$ and $\eta_{j,2}$ associated with the duration, integrated response, and strength of ERK-PP activity. We do this by considering the dynamic behavior, within a time frame of 6 hours, of the MAPK signaling cascade model depicted in Fig. 2.1. We consider two cases: (a) reaction-oriented sensitivity analysis (ROSA), and (b) species-oriented sensitivity analysis (SOSA), as we have explained in Section 3.3. ROSA investigates the importance of reactions in influencing the system response, and SOSA investigates the importance of molecular species in influencing the system response. In each case, we need to set values for the standard deviations $\{\lambda_m^\ddagger,\ m = 1, 2, \ldots, M\}$ of the standard chemical potentials of the activated complexes of the reactions and the standard deviations $\{\lambda_n, n = 1, 2, \ldots, N\}$ of the standard chemical potentials of the molecular species. Due to difficulties in obtaining these values in practice, we assume here that $\lambda_m^\ddagger = \lambda^\ddagger$, for $m = 1, 2, \ldots, M$, and $\lambda_n = \lambda$, for $n = 1, 2, \ldots, N$, and consider $\lambda^\ddagger$, λ as two "user-defined" parameters that quantify the fluctuation levels in biochemical factor values. By following the previous work in Section 3.3, we perform sensitivity analysis with $\lambda^\ddagger, \lambda = 0.1,\ 0.2,\ 0.3,\ 0.4$.[8] Finally, we employ $L = 6{,}000$ Latin hypercube samples in MC and OHA, $S = 4$ regression samples per factor in PA,[9] and a Gauss-Hermite integration of order $Q = 5$ in GHI.

[8] As a consequence of Eq. (2.21), if $Y_n = 0$, for $n = 1, 2, \ldots, N$, then a $\pm\lambda^\ddagger$ variation in the values of $Y_m^\ddagger$ about zero will produce a variation in the nominal values of the rate constants of the m^{th} reaction within the percentage interval $100\left[\exp\{-\lambda_m^\ddagger\} - 1, \exp\{\lambda_m^\ddagger\} - 1\right]\%$. This corresponds to variations in the nominal values of the reaction rate constants within the interval $[-9.52\%, 10.52\%]$, for $\lambda^\ddagger = 0.1$, $[-18.13\%, 22.14\%]$, for $\lambda^\ddagger = 0.2$, $[-25.92\%, 34.99\%]$, for $\lambda^\ddagger = 0.3$, and $[-32.97\%, 49.18\%]$, for $\lambda^\ddagger = 0.4$.

[9] In our simulations, we use $S = 4$ regression points per biochemical factor, located at $-2w, -w, w$, and $2w$, where $w = \lambda^\ddagger$ for ROSA and $w = \lambda$ for SOSA (i.e., we use regression points located at $\pm$ one and two standard deviations from $\mathbf{0}$).

Table 4.1: Number of required system integrations, equations used, and sources of error for each approximation method.

method	**required system integrations**	**ROSA**	**SOSA**	**equations used**	**sources of error**
MC	$2L(J+1)$	264000	288000	Eqs. (4.2)–(4.4)	number of MC samples used
DA	$2J(J+1)+1$	925	1105	Eq. (4.20)	local approximation truncation of Taylor series derivative approximation
PA	$J(J-1)S^2/2$ $+JS+1$	3445	4141	Eq. (4.24)	local approximation truncation of FD-HDMR polynomial approximation polynomial regression
GHI	$2J(J-1)\lfloor Q/2\rfloor^2$ $+2J\lfloor Q/2\rfloor+1$	3445	4141	Eqs. (4.28)– (4.30), (4.33)	local approximation truncation of FD-HDMR Gauss-Hermite integration
OHA	L	6000	6000	Eq. (4.41)	truncation of ANOVA-HDMR Hermite approximation polynomial regression

L: number of Monte Carlo (Latin hypercube) samples

J: number of biochemical factors

S: number of regression samples per factor

Q: order of Gauss-Hermite integration

In Table 4.1, we summarize the number of system integrations and the equations used by each method. For ROSA-based sensitivity analysis ($J = 21$), the number of system integrations required by DA, PA, GHI, and OHA, are respectively only 0.35%, 1.30%, 1.30%, and 2.27% of that required by MC. For SOSA-based sensitivity analysis ($J = 23$), the number of system integrations required by DA, PA, GHI, and OHA, are respectively only 0.38%, 1.44%, 1.44%, and 2.08% of that required by MC.

We list the ROSA results in Tables 4.2–4.4, obtained by the five techniques considered in this chapter and for the four fluctuation levels in the standard chemical potentials of the activated complexes associated with the reactions. The results are given in percentages. For ease of presentation, we have truncated the SESI and JESI values to the closest integers. To reduce the size of the tables, we depict only the results associated with reactions whose truncated SESI or JESI values, estimated by MC, are at least 5%. We consider the SESI and JESI values obtained by MC as the "true" values. By following the previous work in Section 3.1, we classify reactions and molecular species into one of four categories of interest: singularly influential, jointly influential, singularly/jointly influential, and non-influential. We do this by comparing their SESI and JESI values to a 10% threshold. In particular, a reaction is singularly influential if the corresponding SESI value is at least 10% and the JESI value is smaller than 10%, jointly influential if the JESI value is at least 10% and the SESI value is smaller than 10%, singularly/jointly influential if both the SESI and JESI values are at least 10%, and non-influential if both the SESI and JESI values are smaller than 10%. In the remaining of this subsection, we discuss the ROSA results separately for each response characteristic. A similar discussion applies for the SOSA results presented in the Appendix at the end of this chapter.

4.6.1 Duration

Estimation, by MC, of the ROSA-based sensitivity indices associated with the duration of ERK-PP activity produces values that change little with the size $\lambda^{\ddagger}$ of the underlying fluctuations; see Table 4.2. Moreover, the estimated SESI and JESI values indicate that

Table 4.2: ROSA-based SESI and JESI values for the *duration* of ERK-PP activity in the MAPK signaling cascade obtained by the five techniques considered in this chapter. Bold reaction numbers indicate SESI or JESI values, obtained by MC, that are above the 10% threshold.

SESI - DURATION ($\lambda^{\ddagger} = 0.1$)						**JESI - DURATION ($\lambda^{\ddagger} = 0.1$)**					
Reaction	**MC**	**DA**	**PA**	**GHI**	**OHA**	**Reaction**	**MC**	**DA**	**PA**	**GHI**	**OHA**
4	28	28	28	27	28	4	1	0	0	0	0
6	24	26	25	22	25	6	1	0	0	0	0
11	7	7	7	9	8	11	0	0	0	0	0
13	18	18	20	18	19	13	1	0	0	0	0
SESI - DURATION ($\lambda^{\ddagger} = 0.2$)						**JESI - DURATION ($\lambda^{\ddagger} = 0.2$)**					
Reaction	**MC**	**DA**	**PA**	**GHI**	**OHA**	**Reaction**	**MC**	**DA**	**PA**	**GHI**	**OHA**
4	26	27	27	29	27	4	2	1	1	1	1
6	22	25	25	25	23	6	2	1	1	1	1
11	7	7	7	8	8	11	1	0	0	0	0
13	16	17	18	16	17	13	1	1	0	0	0
17	5	5	6	4	5	17	1	1	1	1	1
21	5	5	5	6	5	21	1	1	0	1	1
SESI - DURATION ($\lambda^{\ddagger} = 0.3$)						**JESI - DURATION ($\lambda^{\ddagger} = 0.3$)**					
Reaction	**MC**	**DA**	**PA**	**GHI**	**OHA**	**Reaction**	**MC**	**DA**	**PA**	**GHI**	**OHA**
4	26	26	26	24	26	4	1	2	2	2	2
6	21	24	20	21	21	6	1	2	1	1	1
11	7	6	7	7	8	11	0	1	0	0	0
13	15	16	13	15	15	13	1	1	1	1	1
17	5	4	6	5	5	17	1	2	2	2	1
21	6	5	8	8	6	21	2	2	3	2	1
SESI - DURATION ($\lambda^{\ddagger} = 0.4$)						**JESI - DURATION ($\lambda^{\ddagger} = 0.4$)**					
Reaction	**MC**	**DA**	**PA**	**GHI**	**OHA**	**Reaction**	**MC**	**DA**	**PA**	**GHI**	**OHA**
4	23	24	23	21	25	4	4	3	2	3	3
6	19	22	20	19	21	6	4	3	2	2	2
11	8	6	6	7	9	11	1	1	0	0	0
13	14	15	12	11	15	13	1	2	1	1	1
17	5	4	6	8	5	17	2	3	2	3	1

the duration is primarily affected by reactions 4, 6, and 13 (refer to Fig. 2.1, Table 2.1 and Table 2.2 for identifying these reactions), which exert their influence only singularly (since the SESI values are larger than 10%, whereas the corresponding JESI values are less than 10%). As a matter of fact, all JESI values are negligible, which indicates that the log-duration may be approximately additive,[10] at least within the range of the applied perturbations. It turns out that the SESI's associated with an additive response function can be well estimated by all previous approximation techniques.

From the results depicted in Table 4.2 (and Table 4.5 in the Appendix and the end of this chapter), it is clear that, as compared to MC, the DA, PA, GHI, and OHA consistently provide good approximations to the SESI and JESI values at all fluctuation levels. Moreover, all methods can be used to correctly classify reactions 4, 6, and 13 as being singularly influential.

4.6.2 Integrated Response

Estimation, by MC, of the ROSA-based sensitivity indices associated with the integrated response of ERK-PP activity produces the SESI and JESI values depicted in Table 4.3. These values indicate that the integrated response is primarily influenced by reactions 4 and 6 (refer to Fig. 2.1 and Table 2.1 for identifying these reactions). For small to moderate fluctuations (i.e., for $\lambda^{\ddagger} = 0.1, 0.2$), reactions 4 and 6 influence the integrated response only singularly. However, for large fluctuations (i.e., for $\lambda^{\ddagger} = 0.3, 0.4$), reaction 4 influences the integrated response both singularly and jointly (since both SESI and JESI values are at least 10%), whereas, reaction 6 still influences

[10] Additive response functions do not produce high-order (≥ 2) joint effects and result in zero JESI values [23]. Although a linear response function is additive, the inverse is not necessarily true.

Table 4.3: ROSA-based SESI and JESI values for the *integrated response* of ERK-PP activity in the MAPK signaling cascade obtained by the five techniques considered in this chapter. Bold reaction numbers indicate SESI or JESI values, obtained by MC, that are above the 10% threshold.

SESI - I-RESPONSE ($\lambda^{\ddagger} = 0.1$)						**JESI - I-RESPONSE ($\lambda^{\ddagger} = 0.1$)**					
Reaction	**MC**	**DA**	**PA**	**GHI**	**OHA**	**Reaction**	**MC**	**DA**	**PA**	**GHI**	**OHA**
4	39	39	39	39	39	4	1	0	0	0	0
6	26	27	27	27	27	6	1	0	0	0	0
11	9	10	9	9	9	11	0	0	0	0	0
13	8	8	8	8	8	13	0	0	0	0	0
SESI - I-RESPONSE ($\lambda^{\ddagger} = 0.2$)						**JESI - I-RESPONSE ($\lambda^{\ddagger} = 0.2$)**					
Reaction	**MC**	**DA**	**PA**	**GHI**	**OHA**	**Reaction**	**MC**	**DA**	**PA**	**GHI**	**OHA**
4	37	38	40	40	39	4	5	1	1	2	2
6	25	27	26	26	25	6	4	0	0	1	1
8	5	5	5	5	6	8	2	0	0	1	1
11	7	9	8	8	8	11	1	0	0	0	0
13	6	8	7	7	7	13	1	1	0	0	0
SESI - I-RESPONSE ($\lambda^{\ddagger} = 0.3$)						**JESI - I-RESPONSE ($\lambda^{\ddagger} = 0.3$)**					
Reaction	**MC**	**DA**	**PA**	**GHI**	**OHA**	**Reaction**	**MC**	**DA**	**PA**	**GHI**	**OHA**
4	38	37	43	41	36	**4**	10	2	9	10	11
6	21	26	22	21	21	6	7	1	4	4	6
8	8	4	7	7	7	8	4	0	3	4	5
SESI - I-RESPONSE ($\lambda^{\ddagger} = 0.4$)						**JESI - I-RESPONSE ($\lambda^{\ddagger} = 0.4$)**					
Reaction	**MC**	**DA**	**PA**	**GHI**	**OHA**	**Reaction**	**MC**	**DA**	**PA**	**GHI**	**OHA**
4	36	36	43	40	34	**4**	15	3	18	15	16
6	18	25	16	19	18	6	8	2	7	7	8
8	8	4	8	9	8	8	7	1	6	6	7

the integrated response only singularly.

It is clear from the results depicted in Table 4.3 (and Table 4.6 in the Appendix of this chapter) that all approximation techniques work relatively well for small to moderate fluctuation levels (i.e., for $\lambda^{\ddagger} = 0.1, 0.2$), providing accurate SESI and JESI

values, as compared to the values obtained by MC, and produce correct classification of the reactions. This is true, since the log integrated response may be approximately additive in this case, as indicated by the negligible JESI values. However, for large fluctuations (i.e., for $\lambda^{\ddagger} = 0.3, 0.4$), the log integrated response is not additive anymore and the results obtained by DA deteriorate noticeably,[11] deeming the use of DA inappropriate. As a matter of fact, the DA is not capable of capturing second-order joint effects and the resulting JESI values are very small. If we use the DA results to classify the reactions, then we will erroneously conclude that reaction 4 influences the integrated response only singularly, when $\lambda^{\ddagger} = 0.3, 0.4$.

From the results depicted in Table 4.3 (and Table 4.6 in the Appendix at the end of this chapter), it is clear that, for large fluctuations, GHI and OHA provide good approximations to the sensitivity indices. Moreover, the results indicate that OHA may be a better approximation technique than GHI (e.g., compare the SESI results obtained by GHI and OHA for reaction 4). On the other hand, the results obtained by PA are much better than the results obtained by DA. However, the performance of PA may deteriorate at high fluctuation levels and may become inferior to GHI and OHA (e.g., compare the results obtained by PA, GHI, and OHA for reaction 4). Finally, it is clear that the sensitivity results obtained by GHI and OHA can be used to correctly classify all reactions.

[11] For example, using the JESI results produced by ROSA, the largest differences between the values obtained by DA and MC are 8% and 12% for $\lambda^{\ddagger} = 0.3, 0.4$, respectively.

4.6.3 Strength

Estimation, by MC, of the ROSA-based sensitivity indices associated with the strength of ERK-PP activity produces the SESI and JESI values depicted in Table 4.4. These values indicate that the log strength may be approximately additive when $\lambda^{\ddagger} = 0.1$. However, the log strength becomes nonadditive when $\lambda^{\ddagger} = 0.2, 0.3, 0.4$, since the estimated JESI values are not negligible at these fluctuation levels. Note that, when $\lambda^{\ddagger} = 0.1$, the strength is primarily affected by reactions 4, 6, 8, and 19, which exert their influence only singularly. However, when $\lambda^{\ddagger} = 0.2$, reaction 8 becomes non-influential, reaction 4 influences the strength both singularly and jointly, whereas, reactions 6 and 19 still influence the strength singularly. On the other hand, when $\lambda^{\ddagger} = 0.3, 0.4$, reactions 4 and 6 influence the strength both singularly and jointly, whereas, reaction 8 influences the strength only jointly (since the JESI values are larger than 10%, whereas, the corresponding SESI values are less than 10%).

It is clear from the results depicted in Table 4.4 (and Table 4.7 in the Appendix at the end of this chapter) that all approximation techniques work relatively well when $\lambda^{\ddagger} = 0.1$, producing accurate SESI and JESI values, as compared to the values obtained by MC, and resulting in correct classification of the reactions. However, when $\lambda^{\ddagger} = 0.2, 0.3, 0.4$, DA produces inaccurate results, while the performance of PA and GHI deteriorates noticeably.[12] Once more, OHA consistently provides good results, which can be used to correctly classify the reactions at all fluctuation levels.

[12]For example, considering the JESI results produced by ROSA, the largest differences between the values obtained by DA and MC are 11%, 20% and 23% for $\lambda^{\ddagger} = 0.2, 0.3, 0.4$, respectively. Moreover, the largest differences between the values obtained by PA and MC are 10%, 8% and 5% for $\lambda^{\ddagger} = 0.2, 0.3, 0.4$, respectively. Finally, the largest differences between the values obtained by GHI and MC are 5%, 7% and 5% for $\lambda^{\ddagger} = 0.2, 0.3, 0.4$, respectively.

Table 4.4: ROSA-based SESI and JESI values for the *strength* of ERK-PP activity in the MAPK signaling cascade obtained by the five techniques considered in this chapter. Bold reaction numbers indicate SESI or JESI values, obtained by MC, that are above the 10% threshold.

SESI - STRENGTH ($\lambda^{\ddagger} = 0.1$)						**JESI - STRENGTH ($\lambda^{\ddagger} = 0.1$)**					
Reaction	**MC**	**DA**	**PA**	**GHI**	**OHA**	**Reaction**	**MC**	**DA**	**PA**	**GHI**	**OHA**
4	38	38	36	30	38	4	1	0	0	0	0
6	17	15	15	14	17	6	1	1	0	0	0
8	10	10	9	6	10	8	1	0	0	0	0
11	8	9	9	4	8	11	0	0	0	0	0
19	12	10	12	15	13	19	1	1	0	0	0
SESI - STRENGTH ($\lambda^{\ddagger} = 0.2$)						**JESI - STRENGTH ($\lambda^{\ddagger} = 0.2$)**					
Reaction	**MC**	**DA**	**PA**	**GHI**	**OHA**	**Reaction**	**MC**	**DA**	**PA**	**GHI**	**OHA**
4	32	34	40	39	33	**4**	13	2	3	8	11
6	14	14	14	12	13	6	8	3	1	3	6
8	8	9	11	12	9	8	7	1	1	2	5
17	6	4	6	3	6	17	6	1	1	2	4
19	10	9	11	12	12	19	5	2	1	1	4
SESI - STRENGTH ($\lambda^{\ddagger} = 0.3$)						**JESI - STRENGTH ($\lambda^{\ddagger} = 0.3$)**					
Reaction	**MC**	**DA**	**PA**	**GHI**	**OHA**	**Reaction**	**MC**	**DA**	**PA**	**GHI**	**OHA**
4	31	30	37	37	27	**4**	23	3	22	25	26
6	10	12	12	11	10	**6**	17	5	9	10	15
8	9	8	10	9	8	**8**	11	2	8	9	11
19	6	8	7	6	5	19	5	4	3	3	4
SESI - STRENGTH ($\lambda^{\ddagger} = 0.4$)						**JESI - STRENGTH ($\lambda^{\ddagger} = 0.4$)**					
Reaction	**MC**	**DA**	**PA**	**GHI**	**OHA**	**Reaction**	**MC**	**DA**	**PA**	**GHI**	**OHA**
4	28	25	40	36	26	**4**	28	5	29	27	29
5	2	1	1	0	2	5	6	5	2	2	5
6	10	10	9	11	10	**6**	16	7	11	11	15
8	8	7	8	10	8	**8**	15	3	11	11	14
15	1	0	0	0	2	15	7	5	4	4	7
21	1	0	0	0	1	21	7	4	4	4	8

4.7 Discussion

The previous numerical results demonstrate that, in terms of estimation accuracy, OHA is the best method and DA is the worst, whereas, PA and GHI are in between, with GHI slightly better than PA. To explain why this is so, we must investigate the sources of error introduced by each technique, which we summarize in Table 4.1.

The estimation error produced by the MC approach is mainly due to the finite number L of samples used and decreases slowly as L increases, regardless of the number J of biochemical factors used, at least theoretically.[13]

There are two sources of error associated with DA. First, substantial errors may be introduced due to the fact that DA *locally* approximates the response function by a Taylor series expansion that includes only first- and second-order partial derivatives. Consequently, DA may not produce good estimates of the sensitivity indices under large fluctuations, since a second-order Taylor series approximation of the response function may not be sufficiently accurate over the range of factor values generated by such fluctuations. This is especially true when the response function is nonadditive (as it is the case with the log integrated response and the log strength of ERK-PP in the MAPK example). In such cases, large factor variations may produce substantial joint effects, which cannot be captured by a local second-order Taylor series approximation. This is evident by the fact that, under large fluctuations, the JESI values obtained by DA, associated with the integrated response and strength, are significantly smaller than the ones produced by MC.

[13]Note, however, that to achieve a certain level of accuracy in practice, we may also need to increase L as the number J of biochemical factors increases, due to the exponential growth in the volume of the biochemical factor space when adding extra dimensions, a problem that is usually referred to as "curse of dimensionality."

A second source of error associated with DA is the approximation of the first- and second-order derivatives of the response function by finite-differences.[14] It has been pointed out in [4] that the resulting approximations must be carefully used, since it is difficult to theoretically predict, control, and numerically evaluate their accuracy. Although a number of techniques have been developed to deal with this problem [82], exact evaluation of the response derivatives usually requires simultaneous integration of a set of "sensitivity equations," together with the differential equations governing the underlying molecular concentration dynamics, which turns out to be a very difficult task due to stiffness of the resulting system of differential equations [4].

PA attempts to improve the accuracy of DA by adding high-order derivative terms in the Taylor series expansion of the response function. In addition to the first- and second-order partial derivatives used by the DA, the Taylor series expansion now includes third- and fourth-order partial derivatives that involve only two biochemical factors. Moreover, instead of approximating the derivatives by finite differences, the method avoids such computations by expanding the response function using FD-HDMR, by truncating all components of order ≥ 3, by respectively approximating the first- and second-order FD-HDMR components with second- and fourth-order polynomials, and by estimating the coefficients of these polynomials using regression. Errors are introduced by truncating the FD-HDMR and *locally* approximating the resulting response function by a fourth-order polynomial including only terms involving single and pairs of biochemical factors. As a consequence, PA may not be able to accurately estimate some SESI and JESI values under large fluctuations, since the underlying

[14] In our simulations, we approximate the first- and second-order partial derivatives of the response function by using Eqs. (4.43) and (4.44), with $\Delta = 0.1$.

truncation and polynomial approximation of the response function may not be sufficiently accurate over the range of factor values generated by such fluctuations. Note also that errors can be introduced due to estimating the polynomial coefficients by regression, a situation that cannot be evaluated and controlled easily.[15]

GHI attempts to improve the accuracy of estimating the sensitivity indices by employing the *exact* first- and second-order FD-HDMR components, and numerically calculating the required expectations and variances using Gauss-Hermite integrations. Errors are introduced when truncating the FD-HDMR and evaluating the expectations and variances by one- and two-dimensional Gauss-Hermite integrations. Evaluating and controlling these errors is practically impossible.[16] Truncation of the FD-HDMR essentially corresponds to a *local* approximation of the response function, although this approximation is expected to be more accurate than the Taylor series and polynomial approximations used by DA and PA, respectively. As a consequence, GHI may not be able to accurately estimate some SESI and JESI values under large fluctuations, since the underlying FD-HDMR truncation may not be sufficiently accurate over the range of factor values generated by such fluctuations.

Finally, the errors introduced by OHA are due to approximating the ANOVA-HDMR expansion of the response function by first- and second-order ANOVA-HDMR components, approximating these components with first- and second-order orthonormal Hermite polynomials, and estimating the coefficients of these polynomials using regression. Here, the truncation of high-order ANOVA-HDMR components does not

[15]Using more samples per biochemical factor does not necessarily increase accuracy, especially in polynomial regression [83, 84].

[16]Higher-order Gauss-Hermite integrations do not necessarily produce higher accuracy. This is true only when the integrands are sufficiently smooth, in the sense that they can be well-approximated by polynomials [84].

correspond to a local approximation of the response function, which is why this approximation is more accurate than truncating the FD-HDMR components, as in GHI. In fact, if we consider fluctuation levels at which the higher-order (≥ 3) terms in the variance decomposition scheme given by Eq. (3.1) are negligible, then the higher-order (≥ 3) terms in the ANOVA-HDMR decomposition of the response function will be negligible as well [see Eq. (3.1)]. This is not necessarily true for the higher-order terms in the FD-HDMR decomposition. Therefore, truncating the ANOVA-HDMR decomposition of the response function, as opposed to the FD-HDMR decomposition, is well justified for fluctuation levels at which the response variance is not appreciably influenced by high-order joint effects. Under very large fluctuations, OHA may not accurately estimate the sensitivity indices, since the underlying truncation of ANOVA-HDMR may not be accurate enough due to appreciable high-order (≥ 3) joint effects in the response variance. However, the *global* nature of the approximation methodology employed by OHA, the direct relationship between ANOVA-HDMR and the response variance decomposition scheme given by Eq. (3.1), and the orthonormality properties of the Hermite polynomials, make OHA the most desirable technique for approximating the sensitivity indices, among the techniques considered in this chapter.

Although we have also obtained simulation results for other biochemical reaction systems, we have limited the presentation in this dissertation to the results obtained for the MAPK model depicted in Fig. 2.1. To illustrate various aspects of the approximation techniques and their relative merits, we have chosen the response functions to represent three types of high-dimensional system responses: the log duration, $\ln D$, is approximately additive for the levels of biochemical factor uncertainty considered in

this dissertation, the log integrated response, $\ln I$, is moderately nonadditive, whereas, the log strength, $\ln S$, is highly nonadditive. Based on our experience so far, all our simulation results are consistent with each other and perfectly agree with the theoretical analysis presented in this chapter. We therefore believe that the conclusions based on the MAPK model are general and can be applied to other biochemical reaction systems as well.

It is very important to keep in mind that the four approximation techniques considered in this chapter are based on the assumption that, for most biochemical reaction systems of interest, fluctuations of input biochemical factors will produce only single and second-order joint effects at the output. As a consequence, truncating the HDMR of the response function to a second-order is a natural thing to do. Note that this assumption depends on the particular choice of the biochemical factors used, on how the system response relates to these factors, and on the fluctuation levels used for sensitivity analysis. In general, the approximation methods discussed in this chapter are expected to fail in the presence of high-order (≥ 3) joint effects among biochemical factors. Therefore, it may be necessary in these cases to consider truncated HDMR's that include higher-order basis functions. Extension to this case is straightforward but computationally demanding, since higher-order cases require evaluation of a large number of variance terms in the decomposition scheme given by Eq. (3.1), which can be a tedious thing to do for large biochemical reaction systems.

We should point out here that GHI is based on the methodology proposed in [6,85], which has been effectively used to calculate statistical moments of the responses of high-dimensional mechanical systems subject to randomly fluctuating loads. In this

dissertation, this method has been reformulated to fit the framework of variance-based sensitivity analysis and has been applied to biochemical reaction systems. On the other hand, OHA is based on the methodology proposed in [37, 86–88] for approximating ANOVA-HDMR's using orthonormal basis functions. OHA can also be viewed as a special case of the polynomial chaos expansion (PCE) approach to sensitivity analysis discussed in [89–91], and has been recently employed in [92] for estimating variance-based sensitivity indices in order to learn the topology of a functional network of interactions from given data. To our knowledge, it is the first time that we systematically compare the four approximation techniques and use them to study the sensitivity properties of biochemical reaction systems [93].

To conclude this chapter, we would like to stress the fact that the approximation techniques presented in this chapter have been derived by assuming that the biochemical factors used for sensitivity analysis are statistically independent and that each factor follows a Gaussian distribution. The assumption of statistical independence between the random variables $\{Y_m^{\ddagger}, m = 1, 2, \ldots, M\}$ and $\{Y_n, n = 1, 2, \ldots, N\}$ has been justified in Section 2.4. However, justifying mutual independence within the sets $\{Y_m^{\ddagger}, m = 1, 2, \ldots, M\}$ and $\{Y_n, n = 1, 2, \ldots, N\}$ is a very difficult thing to do. We simply view this assumption as a convenient approximation that allows us to proceed with the sensitivity analysis approaches discussed in this dissertation. Developing variance-based sensitivity analysis for correlated biochemical factors is a challenging problem that needs careful investigation [23,94]. On the other hand, if the biochemical factors follow non-Gaussian distributions, such as uniform, gamma, binomial, etc., the approximation techniques must be appropriately modified to accommodate these dis-

tributions. For example, if each biochemical factor follows a uniform distribution, then we must replace the Gauss-Hermite integration step in GHI by Gauss-Legendre integration [84]. Moreover, if the biochemical factors follow gamma distributions, then we must replace the orthonormal Hermite polynomials in OHA by orthonormal Laguerre polynomials [89, 91].

4.8 Appendix

In Tables 4.5–4.7, we provide the SOSA-based sensitivity analysis results for the three response characteristics (duration, integrated response, and strength) of ERK-PP activity in the MAPK signaling cascade obtained by the five techniques (MC, DA, PA, GHI, and OHA) considered in this chapter and for four fluctuation levels ($\lambda = 0.1, 0.2, 0.3, 0.4$) in the values of the standard chemical potentials associated with the molecular species (refer to Fig. 2.1 and Table 2.3 for identifying these species). The results are given in percentages and have been truncated to the nearest integer. Only results that correspond to SESI or JESI values obtained by MC that are at least 5% are shown. Bold species numbers indicate SESI or JESI values that are at least 10%. According to our discussion in Section 3.1, these species are classified by the variance-based sensitivity analysis method to be *singularly influential* (if the SESI value is at least 10% but the JESI value is below 10%), *jointly influential* (if the JESI value is at least 10% but the SESI value is below 10%), and *singularly/jointly influential* (if both the SESI and JESI values are at least 10%). The remaining molecular species are deemed to be *non-influential*.

Table 4.5: SOSA-based SESI and JESI values for the *duration* of ERK-PP activity in the MAPK signaling cascade obtained by the five techniques considered in this chapter. Bold species numbers indicate SESI or JESI values, obtained by MC, that are above the 10% threshold.

SESI - DURATION ($\lambda = 0.1$)						**JESI - DURATION ($\lambda = 0.1$)**					
Species	**MC**	**DA**	**PA**	**GHI**	**OHA**	**Species**	**MC**	**DA**	**PA**	**GHI**	**OHA**
5	38	37	38	34	38	5	1	0	0	0	0
7	23	25	23	25	23	7	0	0	0	0	0
14	17	17	19	19	18	14	0	0	0	0	0
SESI - DURATION ($\lambda = 0.2$)						**JESI - DURATION ($\lambda = 0.2$)**					
Species	**MC**	**DA**	**PA**	**GHI**	**OHA**	**Species**	**MC**	**DA**	**PA**	**GHI**	**OHA**
5	36	35	36	37	37	5	4	1	1	1	1
7	20	24	22	20	22	7	2	1	1	1	1
14	15	16	17	19	16	14	1	0	0	0	0
SESI - DURATION ($\lambda = 0.3$)						**JESI - DURATION ($\lambda = 0.3$)**					
Species	**MC**	**DA**	**PA**	**GHI**	**OHA**	**Species**	**MC**	**DA**	**PA**	**GHI**	**OHA**
5	36	33	36	33	36	5	3	3	2	3	2
7	20	23	20	21	21	7	1	1	1	1	1
14	15	16	14	14	15	14	1	1	1	1	1
18	5	4	5	6	5	18	1	2	2	2	1
SESI - DURATION ($\lambda = 0.4$)						**JESI - DURATION ($\lambda = 0.4$)**					
Species	**MC**	**DA**	**PA**	**GHI**	**OHA**	**Species**	**MC**	**DA**	**PA**	**GHI**	**OHA**
5	34	31	27	32	33	5	5	4	4	5	5
7	19	21	20	18	19	7	3	2	2	2	3
12	5	4	5	4	6	12	1	1	0	0	1
14	15	15	13	11	15	14	1	1	1	1	1

Table 4.6: SOSA-based SESI and JESI values for the *integrated response* of ERK-PP activity in the MAPK signaling cascade obtained by the five techniques considered in this chapter. Bold species numbers indicate SESI or JESI values, obtained by MC, that are above the 10% threshold.

SESI - I-RESPONSE ($\lambda = 0.1$)						**JESI - I-RESPONSE ($\lambda = 0.1$)**					
Species	**MC**	**DA**	**PA**	**GHI**	**OHA**	**Species**	**MC**	**DA**	**PA**	**GHI**	**OHA**
5	46	47	47	47	47	5	1	0	0	0	0
7	23	23	23	23	23	7	0	0	0	0	0
9	9	9	9	9	9	9	1	0	0	0	0
14	11	12	12	12	12	14	0	0	0	0	0
SESI - I-RESPONSE ($\lambda = 0.2$)						**JESI - I-RESPONSE ($\lambda = 0.2$)**					
Species	**MC**	**DA**	**PA**	**GHI**	**OHA**	**Species**	**MC**	**DA**	**PA**	**GHI**	**OHA**
5	47	46	50	49	46	5	7	1	2	5	5
7	19	23	21	20	21	7	4	0	1	2	2
9	8	9	9	8	9	9	3	0	1	2	3
14	8	12	9	9	9	14	1	0	0	0	0
SESI - I-RESPONSE ($\lambda = 0.3$)						**JESI - I-RESPONSE ($\lambda = 0.3$)**					
Species	**MC**	**DA**	**PA**	**GHI**	**OHA**	**Species**	**MC**	**DA**	**PA**	**GHI**	**OHA**
5	47	45	50	52	44	**5**	14	2	16	14	15
7	16	23	15	15	16	7	7	1	6	5	6
9	9	9	9	8	9	9	5	1	6	5	7
SESI - I-RESPONSE ($\lambda = 0.4$)						**JESI - I-RESPONSE ($\lambda = 0.4$)**					
Species	**MC**	**DA**	**PA**	**GHI**	**OHA**	**Species**	**MC**	**DA**	**PA**	**GHI**	**OHA**
5	45	44	45	48	46	**5**	16	3	22	17	15
7	15	22	13	14	15	7	8	1	8	7	7
9	9	8	9	10	9	9	7	1	8	7	7

Table 4.7: SOSA-based SESI and JESI values for the *strength* of ERK-PP activity in the MAPK signaling cascade obtained by the five techniques considered in this chapter. Bold species numbers indicate SESI or JESI values, obtained by MC, that are above the 10% threshold.

SESI - STRENGTH ($\lambda = 0.1$)						**JESI - STRENGTH ($\lambda = 0.1$)**					
Species	**MC**	**DA**	**PA**	**GHI**	**OHA**	**Species**	**MC**	**DA**	**PA**	**GHI**	**OHA**
5	40	41	40	38	41	5	1	0	0	0	0
7	13	11	14	8	13	7	1	0	0	0	0
9	26	26	27	29	26	9	1	0	0	0	0
17	5	6	5	5	6	17	0	0	0	0	0
21	6	6	5	8	6	21	0	0	0	0	0
SESI - STRENGTH ($\lambda = 0.2$)						**JESI - STRENGTH ($\lambda = 0.2$)**					
Species	**MC**	**DA**	**PA**	**GHI**	**OHA**	**Species**	**MC**	**DA**	**PA**	**GHI**	**OHA**
5	40	38	47	46	35	**5**	18	2	10	18	17
7	10	10	11	11	10	7	9	1	4	6	7
9	15	24	17	16	17	9	9	1	4	6	9
SESI - STRENGTH ($\lambda = 0.3$)						**JESI - STRENGTH ($\lambda = 0.3$)**					
Species	**MC**	**DA**	**PA**	**GHI**	**OHA**	**Species**	**MC**	**DA**	**PA**	**GHI**	**OHA**
5	41	35	44	49	34	**5**	27	3	30	28	29
7	8	9	7	7	8	**7**	15	1	11	9	13
9	10	22	10	9	10	9	9	1	10	9	9
22	1	0	0	0	1	22	6	5	4	4	7
SESI - STRENGTH ($\lambda = 0.4$)						**JESI - STRENGTH ($\lambda = 0.4$)**					
Species	**MC**	**DA**	**PA**	**GHI**	**OHA**	**Species**	**MC**	**DA**	**PA**	**GHI**	**OHA**
5	40	31	40	41	39	**5**	26	5	35	29	26
7	8	8	7	8	8	**7**	13	2	12	11	13
9	9	20	8	10	9	**9**	11	2	11	10	11
22	2	0	1	1	2	22	6	8	5	5	7

Chapter 5

Reducing Experimental Variability

Applying the previously proposed variance-based sensitivity analysis technique (and almost any other sensitivity analysis technique available in the literature) requires specification of the nominal values of underlying kinetic parameters [32]. It turns out that different nominal values may produce different sensitivity results (see our example in Section 5.6 and the Appendix at the end of this chapter), which can be very problematic in applications of the method to systems biology problems. For the sensitivity results to be biologically relevant, the nominal values must be close to their "true" values. However, the true parameter values of real biochemical reaction systems are hardly known. Most often, nominal kinetic parameter values are estimated from measurements reported by various laboratories using different experimental methods and conditions. It is therefore uncertain whether a particular set of nominal parameters closely approximates the biological system at hand [45, 95]. Even if nominal parameter values are obtained under the same experimental conditions (which is usually done by a collective fit approach based on minimizing the discrepancy between obtained exper-

imental data and predicted system responses), there is a high degree of uncertainty on whether these values are the "true" system values. This is due to the "sloppiness" of biochemical reaction systems (different combinations of parameter values may lead to equally acceptable concentration dynamics) [96, 97], appreciable errors typical to experimental biology [98], as well as substantial differences between the in-vitro and in-vivo kinetic properties of molecular species [99].

In fact, biological data are always subject to unpredictable biological and experimental variabilities. When dealing with such data, it is important to distinguish between these two sources of uncertainty, so that experimental protocols and computational analysis can be appropriately designed and conclusions can be accurately drawn [39, 100]. For example, when using ROSA and SOSA in systems-based drug design applications, we are only interested in assessing how biological variability influences cellular behavior, since experimental variability does not exist in living cells and cannot be pharmacologically controlled. If experimental variability is prominent, then its effect should be minimized in order to improve the accuracy of the sensitivity analysis method used.

Motivated by this need, we propose in this chapter a variance-based approach to sensitivity analysis that explicitly considers experimental variability and effectively reduces its effects on the obtained results. More specifically, we introduce a new set of sensitivity indices, which we refer to as noise-reduced variance-based sensitivity indices (NR-VSI's), that account for the uncertainty in nominal reaction rate values due to experimental variability. Although these indices are a simple extension of the variance-based sensitivity indices considered in Chapter 3, they are less sensitive to the choice

of nominal rate values used and lead to a powerful sensitivity analysis methodology that can effectively accommodate different levels of experimental variability.

Variance-based sensitivity indices can be estimated by Monte Carlo sampling. This is a powerful yet computationally demanding numerical approach, since it requires evaluation of the response function of a biochemical reaction system at a large number of kinetic parameter values. To reduce computational effort, many approximation techniques have been proposed in the literature [75–79, 90, 92, 93]. Most techniques replace the actual response function by an approximate one that is much easier to compute, and estimate the sensitivity indices either analytically or by Monte Carlo sampling using the approximate function. However, a major limitation of this approach is that it does not work well for large biochemical reaction systems, since it cannot capture the effects of high-order biochemical factor interactions on high-dimensional nonlinear system responses [76, 101].

One way to address this problem is to use a pre-screening technique to identify non-influential biochemical factors, whose contributions to the system response variance are negligible, then fix their values to the corresponding nominal values, or group them together and treat them as one factor [102]. This approach will reduce the dimensionality of the sensitivity analysis problem at hand and make estimation of sensitivity indices easier to handle. Currently available pre-screening techniques employ elementary effect measures [103, 104], or integrals of squared partial derivatives [105, 106]. Although these methods work well for the analytical examples provided in the literature, they are unreliable when used for sensitivity analysis of large biochemical reaction systems [101]. Both methods are prone to large errors, since they require evaluation

of partial derivative of the response function with respect to each biochemical factor at different points in the parameter space, which cannot be done analytically. Partial derivatives are usually approximated by finite differences, whose accuracy is very difficult to assess and control [4]. When evaluating elementary effect measures or integrals of squared partial derivatives, numerical errors may accumulate and propagate, and in some cases grow exponentially or oscillate [107], causing the calculated pre-screening measures to appreciably deviate from their actual values, thus leading to unreliable sensitivity analysis results.

On the other hand, in real biological systems, substantial high-order interactions normally exist only among those system factors that already demonstrate substantial low-order interactions [108]. This motivates us to assume that, if the singular and joint contribution to the system response variance of a biochemical factor with another factor is negligible, it is unlikely that this factor will contribute to the system response variance by interacting with multiple other factors, in which case the biochemical factor will be non-influential. As a consequence of this assumption, we can ignore high-order (≥ 3) factor contributions to the response variance, which are difficult to compute, and use an approximate second-order sensitivity analysis method, such as one of the techniques discussed in Chapter 4, to pre-screen non-influential factors. As a matter of fact, by employing an orthonormal Hermite polynomial approximation (OHA) of the response function, we can construct a second-order variance-based sensitivity analysis technique for accurately pre-screening non-influential biochemical factors. This leads to an attractive dimensionality reduction approach that can be used to effectively reduce the computational effort required by estimating the NR-VSI's using Monte Carlo sampling.

5.1 Probabilistic Modeling of Rate Constants Revisited

In Section 2.4, we have treated the rate constants k_{2m-1} and k_{2m} as random variables K_{2m-1} and K_{2m}, given by the following Eyring-Polanyi equations:

$$K_{2m-1} = \frac{k_B T}{h} \frac{C_m^{\ddagger}}{\prod_{n=1}^{N} C_n^{\nu_{nm}}} \qquad \text{and} \qquad K_{2m} = \frac{k_B T}{h} \frac{C_m^{\ddagger}}{\prod_{n=1}^{N} C_n^{\nu'_{nm}}}, \tag{5.1}$$

where k_B is the Boltzmann constant ($k_B = 1.3806504 \times 10^{-23}\text{JK}^{-1}$), T is the system temperature, h is the Planck constant ($h = 6.62606885 \times 10^{-34}\text{Js}$), $C_m^{\ddagger}$ is the (random) capacity of the activated complex associated with the m^{th} reaction, and C_n is the (random) capacity of the n^{th} molecular species. The capacities are defined by

$$C_m^{\ddagger} := x_{\text{total}} \exp\left\{-\frac{M_m^{\ddagger 0}}{k_B T}\right\} \qquad \text{and} \qquad C_n := x_{\text{total}} \exp\left\{-\frac{M_n^0}{k_B T}\right\}, \tag{5.2}$$

where $M_m^{\ddagger 0}$, M_n^0 are the (random) standard chemical potentials of the m^{th} activated complex and the n^{th} molecular species, respectively, given by

$$M_m^{\ddagger 0} = \mu_m^{\ddagger 0} + k_B T Y_m^{\ddagger} \qquad \text{and} \qquad M_n^0 = \mu_n^0 + k_B T Y_n. \tag{5.3}$$

In Eq. (5.3), $\mu_m^{\ddagger 0}$ and μ_n^0 are the nominal standard chemical potential values associated with the m^{th} reaction and the n^{th} molecular species, respectively, whereas, $Y_m^{\ddagger}$ and Y_n are zero-mean Gaussian random variables with standard deviations $\lambda_m^{\ddagger}$ and λ_n, respectively. These random variables account for fluctuations in the standard chemical potentials about their nominal values caused by unpredictable *biological* variability. Following our previous approach, we will again assume here that the random variables $\{Y_m^{\ddagger}, m = 1, 2, \ldots, M\}$ and $\{Y_n, n = 1, 2, \ldots, N\}$ are statistically independent and that $\lambda_m^{\ddagger} = \lambda^{\ddagger}$, for every $m = 1, 2, \ldots, M$, whereas, $\lambda_n = \lambda$, for every $n = 1, 2, \ldots, N$.

As a consequence of Eqs. (5.1)–(5.3), we have that

$$\widetilde{\mathbf{K}} = \widetilde{\mathbf{k}}^{(0)} + \mathbb{A}\mathbf{Y}. \tag{5.4}$$

In this equation, $\widetilde{\mathbf{K}}$ is a $2M \times 1$ random vector with elements $\{\widetilde{K}_{2m-1} := \ln K_{2m-1}$, $\widetilde{K}_{2m} := \ln K_{2m}, m = 1, 2, \ldots, M\}$, and $\widetilde{\mathbf{k}}^{(0)}$ is a $2M \times 1$ vector with elements $\{\widetilde{k}^{(0)}_{2m-1} := \ln k^{(0)}_{2m-1}$, $\widetilde{k}^{(0)}_{2m} := \ln k^{(0)}_{2m}, m = 1, 2, \ldots, M\}$, where $k^{(0)}_{2m-1}$ and $k^{(0)}_{2m}$ are the nominal values of the rate constants given in Eq. (2.14). In addition, $\mathbb{A}$ is a $2M \times (M + N)$ matrix, given by

$$\mathbb{A} = \begin{bmatrix} -1 & 0 & \cdots & 0 & \nu_{11} & \nu_{21} & \cdots & \nu_{N1} \\ -1 & 0 & \cdots & 0 & \nu'_{11} & \nu'_{21} & \cdots & \nu'_{N1} \\ 0 & -1 & \cdots & 0 & \nu_{12} & \nu_{22} & \cdots & \nu_{N2} \\ 0 & -1 & \cdots & 0 & \nu'_{12} & \nu'_{22} & \cdots & \nu'_{N2} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 & \nu_{1M} & \nu_{2M} & \cdots & \nu_{NM} \\ 0 & 0 & \cdots & -1 & \nu'_{1M} & \nu'_{2M} & \cdots & \nu'_{NM} \end{bmatrix}, \tag{5.5}$$

whereas, $\mathbf{Y}$ is an $(M+N)\times 1$ random vector with elements $\{Y^{\ddagger}_m, m = 1, 2, \ldots, M, Y_n, n = 1, 2, \ldots, N\}$. Eq. (5.4) provides a probabilistic model for perturbing the forward and reverse reaction rate constants of a biochemical reaction system about their nominal values. In addition to being derived from basic biophysical principles, it guarantees that the perturbed reaction rate constants automatically satisfy the necessary Wegscheider conditions imposed by thermodynamics, which are given by Eq. (2.25), provided that these conditions are satisfied by the nominal reaction rate constants. As a consequence, Eq. (5.4) leads to a thermodynamically consistent method for sensitivity analysis of biochemical reaction systems [32].

Determining the true nominal values $\{k^{(0)}_{2m-1}, k^{(0)}_{2m}, m = 1, 2, \ldots, M\}$ for the rate constants of a given biochemical reaction system is a difficult task, usually addressed

by directly measuring these values or by collectively fitting them to experimental data using statistical inference techniques. Due to variations in experimental conditions and biochemical methods used, directly measuring the kinetic parameters usually results in reporting different nominal values in the literature. As we discussed above, this can be problematic for sensitivity analysis, since the results may appreciably depend on the particular set of nominal values used. On the other hand, a statistical inference technique for parameter estimation does not usually produce a unique set of nominal values but a statistical distribution over a range of possible values. This is true, for example, in Bayesian inference, which characterizes parameter estimates by a probability distribution, known as posterior [109].

The previous discussion suggests that, in many practical situations, we must treat the nominal reaction rate constants as random variables $\{K_{2m-1}^{(0)}, K_{2m}^{(0)}, m = 1, 2, \ldots, M\}$. This allows us to quantify the uncertainty about the true nominal values of the rate constants by means of a joint probability distribution and obtain representative values by sampling this distribution. In the following, we suggest a particular probability distribution for the nominal reaction rate constants, derived from the Eyring-Polanyi equations. Although other choices are possible, such as the posterior distribution obtained by a Bayesian parameter estimation approach, our choice is simple and sufficiently illustrates the sensitivity analysis methodology we discuss in this chapter.

In the following, we will assume that the logarithms $\{\widetilde{K}_{2m-1}^{(0)} := \ln K_{2m-1}^{(0)}, \widetilde{K}_{2m}^{(0)} := \ln K_{2m}^{(0)}, m = 1, 2, \ldots, M\}$ of the nominal reaction rate constants follow a multivariate Gaussian distribution with mean values $\{\mu_{2m-1} := \mathrm{E}\big[\ln K_{2m-1}^{(0)}\big],\ \mu_{2m} := \mathrm{E}\big[\ln K_{2m}^{(0)}\big], m = 1, 2, \ldots, M\}$. Our objective is to find a *column orthogonal* $2M \times P$ matrix $\mathbb{V}$ (i.e.,

a $2M \times P$ matrix $\mathbb{V}$ such that $\mathbb{V}^T\mathbb{V} = \mathbb{I}_{P\times P}$, where $\mathbb{I}_{P\times P}$ is the $P \times P$ identity matrix), such that

$$\widetilde{\mathbf{K}}^{(0)} = \boldsymbol{\mu} + \mathbb{V}\mathbf{Z}, \tag{5.6}$$

where $\widetilde{\mathbf{K}}^{(0)}$ is a $2M \times 1$ random vector with elements $\{\widetilde{K}^{(0)}_{2m-1}, \widetilde{K}^{(0)}_{2m}, m = 1, 2, \ldots, M\}$, $\boldsymbol{\mu}$ is a $2M \times 1$ vector with elements $\{\mu_{2m-1}, \mu_{2m}, m = 1, 2, \ldots, M\}$, and $\mathbf{Z}$ is a $P \times 1$ random vector with independently and identically distributed zero-mean Gaussian random elements $\{Z_p, p = 1, 2, \ldots, P\}$ of variance σ^2. Then, $\widetilde{\mathbf{K}}^{(0)}$ will be a multivariate Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\sigma^2\mathbb{V}\mathbb{V}^T$. Note that, due to the column orthogonality of matrix $\mathbb{V}$, there will be a one-to-one correspondence between $\widetilde{\mathbf{K}}^{(0)}$ and $\mathbf{Z}$, since $\mathbf{Z} = \mathbb{V}^T(\widetilde{\mathbf{K}}^{(0)} - \boldsymbol{\mu})$ in this case. This allows us to efficiently sample the multivariate Gaussian distribution of $\widetilde{\mathbf{K}}^{(0)}$ in a Monte Carlo setting by independently sampling the univariate Gaussian distributions of the elements in $\mathbf{Z}$.[17]

To solve the previous problem, note that the nominal reaction rate constants $\{K^{(0)}_{2m-1}, K^{(0)}_{2m}, m = 1, 2, \ldots, M\}$ must also satisfy the Eyring-Polanyi equations, given by Eqs. (5.1)–(5.3). As a consequence,

$$\widetilde{\mathbf{K}}^{(0)} = \mathbf{x_T} + \mathbb{A}\mathbf{M}. \tag{5.7}$$

[17]If the linear transformation given by Eq. (5.6) is not one-to-one, then an infinite number of $\mathbf{Z}$ values will be mapped to the same value of $\widetilde{\mathbf{K}}^{(0)}$. This produces inefficient sampling of $\widetilde{\mathbf{K}}^{(0)}$, since a lot of effort may be wasted by repeatedly sampling the same value of $\widetilde{\mathbf{K}}^{(0)}$, and can result in a much slower converge of the Monte Carlo algorithm used to estimate the sensitivity indices discussed in this chapter.

In this equation, $\mathbf{x_T}$ is the $2M \times 1$ vector given by

$$\mathbf{x_T} = \begin{bmatrix} \ln \frac{k_B T}{h} + \ln x_{\text{total}} - \sum_{n=1}^{N} \nu_{n1} x_{\text{total}} \\ \ln \frac{k_B T}{h} + \ln x_{\text{total}} - \sum_{n=1}^{N} \nu'_{n1} x_{\text{total}} \\ \ln \frac{k_B T}{h} + \ln x_{\text{total}} - \sum_{n=1}^{N} \nu_{n2} x_{\text{total}} \\ \ln \frac{k_B T}{h} + \ln x_{\text{total}} - \sum_{n=1}^{N} \nu'_{n2} x_{\text{total}} \\ \vdots \\ \ln \frac{k_B T}{h} + \ln x_{\text{total}} - \sum_{n=1}^{N} \nu_{nM} x_{\text{total}} \\ \ln \frac{k_B T}{h} + \ln x_{\text{total}} - \sum_{n=1}^{N} \nu'_{nM} x_{\text{total}} \end{bmatrix},$$

$\mathbf{M}$ is an $(M+N) \times 1$ random vector with elements $\{M_m^{\ddagger 0}/(k_B T),\ m = 1, 2, \ldots, M,$ $M_n^0/(k_B T),\ n = 1, 2, \ldots, N\}$, and $\mathbb{A}$ is the $2M \times (M+N)$ matrix given by Eq. (5.5). Eq. (5.7) shows that the nominal rate constant values are spanned by the columns of matrix $\mathbb{A}$, which are not orthonormal. On the other hand, and according to Eq. (5.6), we want these values to also be spanned by the columns of matrix $\mathbb{V}$, which must be orthonormal. We can accomplish this by setting the columns of matrix $\mathbb{V}$ equal to the basis vectors of the column space of matrix $\mathbb{A}$.

As a consequence of Eqs. (5.4) and (5.6), we have that

$$\widetilde{\mathbf{K}} = \boldsymbol{\mu} + \mathbb{A}\mathbf{Y} + \mathbb{V}\mathbf{Z}, \tag{5.8}$$

where $\mathbb{A}$ is given by Eq. (5.5) and the columns of $\mathbb{V}$ form an orthonormal basis for the columns of $\mathbb{A}$. Eq. (5.8) provides an extension of the probabilistic model for the rate constants of a biochemical reaction system proposed in Section 2.4, by incorporating the effects of both *biological* and *experimental* variability on these constants by means of two well-defined random terms, namely $\mathbb{A}\mathbf{Y}$ and $\mathbb{V}\mathbf{Z}$, respectively. The "level" of biological variability is quantified by the standard deviations $\lambda^{\ddagger}$ (for reactions) and λ (for molecular species), associated with the zero-mean Gaussian random variables $Y_m^{\ddagger}$

and Y_n, respectively, whereas, the "level" of experimental variability is quantified by the standard deviation σ, associated with the zero-mean Gaussian random variable Z_p.

To use Eq. (5.8) for sensitivity analysis, we must know the mean values $\{\mu_{2m-1}, \mu_{2m}, m = 1, 2, \ldots, M\}$ of the nominal log rate constants. Moreover, we must make sure that the exponentials of these values satisfy the necessary Wegscheider conditions implied by thermodynamics; i.e.,

$$\prod_{m=1}^{M} \left[\frac{\exp\{\mu_{2m-1}\}}{\exp\{\mu_{2m}\}} \right]^{b_m} = 1, \quad \text{for all} \;\; \mathbf{b} \in \text{null}(\mathbb{S}). \tag{5.9}$$

This guarantees that $\{K^{(0)}_{2m-1}, K^{(0)}_{2m}, m = 1, 2, \ldots, M\}$, and therefore $\{K_{2m-1}, K_{2m}, m = 1, 2, \ldots, M\}$, satisfy the Wegscheider conditions. In practice, we can estimate $\{\mu_{2m-1}, \mu_{2m}, m = 1, 2, \ldots, M\}$ by averaging published values for the rate constants, if available, or by Monte Carlo sampling of the posterior distribution, if a Bayesian inference approach is used for parameter estimation. If the values $\{\mu_{2m-1}, \mu_{2m}, m = 1, 2, \ldots, M\}$ do not satisfy Eq. (5.9), then we can use an orthogonal projection technique, discussed in [110], to map them to thermodynamically feasible values.

5.2 Noise-Reduced Sensitivity Indices

To enhance the predictive power and robustness of sensitivity analysis, we need to separate biological and experimental variability and reduce the effects of experimental variability on the results of such analysis. These are important tasks, since, in a systems biology setting, the main objective of sensitivity analysis is to assess how biological variability affects cellular behavior. Eq. (5.8) derived in the previous section provides a probabilistic model for the rate constants of a biochemical reaction system

that effectively separates biological and experimental variability. In this section, we use this model to develop a variance-based sensitivity analysis approach that exploits the effects of biological variability by appropriately averaging experimental variability out of the problem.

Similar to Section 3.1, we use $U_1, U_2, \ldots, U_J$ to denote the random variables $Y^{\ddagger}$ and Y associated with biological variability. We consider two cases, namely $J = M$ and $U_j = Y_j^{\ddagger}$, for $j = 1, 2, \ldots, M$, as well as $J = N$ and $U_j = Y_j$, for $j = 1, 2, \ldots, N$. In the first case, the objective is to perform *reaction-oriented sensitivity analysis* (ROSA) in order to investigate the importance of reactions in influencing the system response. In the second case, the objective is to perform *species-oriented sensitivity analysis* (SOSA) in order to investigate the importance of molecular species in influencing the system response. Recall that we use $Z_1, Z_2, \ldots, Z_P$ to denote the random variables associated with experimental variability. We will be referring to $\mathbf{U} = \{U_1, U_2, \ldots, U_J\}$ as "biochemical factors" and to $\mathbf{Z} = \{Z_1, Z_2, \ldots, Z_P\}$ as "noise factors."

As explained in Section 2.3, the response R of a biochemical reaction system is calculated by solving the system of differential equations given by Eq. (2.2). Moreover, and as a consequence of Eq. (5.8), R is a function of both $\mathbf{U}$ and $\mathbf{Z}$. We can assess the sensitivity properties of the biochemical reaction system due to biological variability by setting $\mathbf{Z} = \mathbf{z}$, for some known value $\mathbf{z}$, and by randomly perturbing $\mathbf{U}$. In this case, the variance $V(\mathbf{z}) := \mathrm{Var}_{\mathbf{U}}\left[R(\mathbf{U}, \mathbf{z})\right]$ of the system response satisfies [25, 26, 74]:

$$V(\mathbf{z}) = \sum_{j=1}^{J} V_j(\mathbf{z}) + \sum_{j=1}^{J-1} \sum_{j'=j+1}^{J} V_{jj'}(\mathbf{z}) + \cdots + V_{12\cdots J}(\mathbf{z}), \tag{5.10}$$

where

$$V_j(\mathbf{z}) := \mathrm{Var}_{U_j}\big[\mathrm{E}_{\mathbf{U}_{(j)}}\big[R(\mathbf{U},\mathbf{z}) \mid U_j\big]\big],$$
$$V_{jj'}(\mathbf{z}) := \mathrm{Var}_{U_j,U_{j'}}\big[\mathrm{E}_{\mathbf{U}_{(j,j')}}\big[R(\mathbf{U},\mathbf{z}) \mid U_j, U_{j'}\big]\big] - V_j(\mathbf{z}) - V_{j'}(\mathbf{z}), \tag{5.11}$$

with similar expressions for the remaining terms. Here, $\mathbf{U}_{(j)}$ denotes all elements in $\mathbf{U}$ excluding U_j, whereas, $\mathbf{U}_{(j,j')}$ denotes all elements in $\mathbf{U}$ excluding U_j and $U_{j'}$. If the biochemical factors $U_1, U_2, \ldots, U_J$ are statistically independent (which we have assumed here to be true), then each term on the right-hand-side of Eq. (5.10) will be nonnegative [the $V_j(\mathbf{z})$ terms are always nonnegative]. This equation provides a decomposition of the system response variance $V(\mathbf{z})$ into individual terms $V_1(\mathbf{z}), V_2(\mathbf{z}), \ldots, V_{12}(\mathbf{z}), \ldots$. By following the discussion in Section 3.1, it turns out that $V_j(\mathbf{z})$ quantifies the average reduction in response variance obtained by keeping the j^{th} biochemical factor fixed. As a consequence, we use $V_j(\mathbf{z})$ to measure the *singular* influence of the j^{th} biochemical factor U_j on the system response. Moreover, the term $V_{jj'}(\mathbf{z})$ quantifies the average reduction in the total response variance due to jointly fixing the two biochemical factors U_j and $U_{j'}$, which is not accounted for by summing the average reductions obtained by separately fixing these factors. Therefore, we use $V_{jj'}(\mathbf{z})$ to measure the *joint* influence of the biochemical factors U_j and $U_{j'}$ on the system response. Finally, higher-order terms in Eq. (5.10) quantify the joint influence of three or more biochemical factors on the system response.

By using the decomposition scheme in Eq. (5.10), we have developed in Chapter 3 a powerful methodology for sensitivity analysis of biochemical reaction systems. The method is based on calculating the *single-effect sensitivity indices* (SESI's), defined by

$$\sigma_j(\mathbf{z}) := \frac{V_j(\mathbf{z})}{V(\mathbf{z})} = \frac{\mathrm{Var}_{U_j}\big[\mathrm{E}_{\mathbf{U}_{(j)}}\big[R(\mathbf{U},\mathbf{z}) \mid U_j\big]\big]}{\mathrm{Var}_{\mathbf{U}}\big[R(\mathbf{U},\mathbf{z})\big]}, \quad j = 1, 2, \ldots, J, \tag{5.12}$$

the *total-effect sensitivity indices* (TESI's), defined by

$$\begin{aligned}\tau_j(\mathbf{z}) &:= \frac{V_j(\mathbf{z}) + \sum_{j'=1,j'\neq j}^{J} V_{jj'}(\mathbf{z}) + \cdots + V_{12\cdots J}(\mathbf{z})}{V(\mathbf{z})} \\ &= \frac{\mathrm{E}_{\mathbf{U}_{(j)}}\big[\mathrm{Var}_{U_j}\big[R(\mathbf{U},\mathbf{z}) \mid \mathbf{U}_{(j)}\big]\big]}{\mathrm{Var}_{\mathbf{U}}\big[R(\mathbf{U},\mathbf{z})\big]}, \quad j = 1, 2, \ldots, J, \end{aligned} \tag{5.13}$$

and the *joint-effect sensitivity indices* (JESI's), defined by

$$\eta_j(\mathbf{z}) := \tau_j(\mathbf{z}) - \sigma_j(\mathbf{z}), \quad j = 1, 2, \ldots, J. \tag{5.14}$$

The SESI $\sigma_j(\mathbf{z})$ quantifies the fractional singular contribution of the j^{th} biochemical factor to the response variance $V(\mathbf{z})$, the JESI $\eta_j(\mathbf{z})$ quantifies the fractional contribution of the j^{th} biochemical factor to $V(\mathbf{z})$ jointly with one or more other factors, whereas, the TESI $\tau_j(\mathbf{z})$ quantifies the fractional total (singular and joint) contribution of the j^{th} biochemical factor to $V(\mathbf{z})$.

Unfortunately, sensitivity analysis based on Eqs. (5.10)–(5.14) depends on the particular choice for the value of the noise factor $\mathbf{Z}$. As a matter of fact, since $\mathbf{Z}$ is a random vector, $V(\mathbf{Z})$ is a random variable and the same is true for the sensitivity indices. If the value $\mathbf{z}_t$ of $\mathbf{Z}$ that corresponds to the true nominal values of the reaction rate constants were known, then we could use the system response variance $V(\mathbf{z}_t)$ and the associated indices $\sigma_j(\mathbf{z}_t)$, $\tau_j(\mathbf{z}_t)$, and $\eta_j(\mathbf{z}_t)$ for sensitivity analysis. However, this value is unknown and sensitivity analysis based on $V(\mathbf{z}_t)$ is not possible.

Since the main objective of sensitivity analysis is to determine the biochemical factors that influence the system response, we must base our approach on that portion of the response variance $\mathrm{Var}_{\mathbf{U},\mathbf{Z}}\big[R(\mathbf{U},\mathbf{Z})\big]$ that encapsulates only the effects of biological variability. Note that the variance $\mathrm{Var}_{\mathbf{Z}}\big[\mathrm{E}_{\mathbf{U}}\big[R(\mathbf{U},\mathbf{Z})\big]\big]$ of the mean system response $\mathrm{E}_{\mathbf{U}}\big[R(\mathbf{U},\mathbf{Z})\big]$, averaged over the effects of biological variability, encapsulates

only experimental variability. As a consequence, the difference $\mathrm{Var}_{\mathbf{U},\mathbf{Z}}\big[R(\mathbf{U},\mathbf{Z})\big] - \mathrm{Var}_{\mathbf{Z}}\big[\mathrm{E}_{\mathbf{U}}\big[R(\mathbf{U},\mathbf{Z})\big]\big]$ is the portion of the system response variance due to biological variability. By using the well-known variance decomposition formula $\mathrm{Var}_Y\big[Y\big] = \mathrm{Var}_X\big[\mathrm{E}_Y\big[Y|X\big]\big] + \mathrm{E}_X\big[\mathrm{Var}_Y\big[Y|X\big]\big]$, we have that

$$\mathrm{Var}_{\mathbf{U},\mathbf{Z}}\big[R(\mathbf{U},\mathbf{Z})\big] - \mathrm{Var}_{\mathbf{Z}}[\mathrm{E}_{\mathbf{U}}\big[R(\mathbf{U},\mathbf{Z})\big]\big] = \mathrm{E}_{\mathbf{Z}}\big[\mathrm{Var}_{\mathbf{U}}\big[R(\mathbf{U},\mathbf{Z})\big]\big]. \tag{5.15}$$

This suggests that we develop a sensitivity analysis approach based on the mean response variance $\overline{V} := \mathrm{E}\big[V(\mathbf{Z})\big] = \mathrm{E}_{\mathbf{Z}}\big[\mathrm{Var}_{\mathbf{U}}\big[R(\mathbf{U},\mathbf{Z})\big]\big]$.

By taking expectations on both sides of Eq. (5.10), and by using Eq. (5.11), we have that

$$\overline{V} = \sum_{j=1}^{J} \overline{V}_j + \sum_{j=1}^{J-1} \sum_{j'=j+1}^{J} \overline{V}_{jj'} + \cdots + \overline{V}_{12\cdots J}, \tag{5.16}$$

where

$$\begin{aligned}
&\overline{V} := \mathrm{E}_{\mathbf{Z}}\big[\mathrm{Var}_{\mathbf{U}}\big[R(\mathbf{U},\mathbf{Z}) \mid \mathbf{Z}\big]\big], \\
&\overline{V}_j := \mathrm{E}_{\mathbf{Z}}\big[\mathrm{Var}_{U_j}\big[\mathrm{E}_{\mathbf{U}_{(j)}}\big[R(\mathbf{U},\mathbf{Z}) \mid U_j, \mathbf{Z}\big]\big]\big], \\
&\overline{V}_{jj'} := \mathrm{E}_{\mathbf{Z}}\big[\mathrm{Var}_{U_j,U_{j'}}\big[\mathrm{E}_{\mathbf{U}_{(j,j')}}\big[R(\mathbf{U},\mathbf{Z}) \mid U_j, U_{j'}, \mathbf{Z}\big]\big]\big] - \overline{V}_j - \overline{V}_{j'},
\end{aligned} \tag{5.17}$$

with similar expressions for the remaining terms. This variance decomposition scheme leads to the following sensitivity indices [compare with Eqs. (5.12)–(5.14)]:

$$\begin{aligned}
\overline{\sigma}_j &:= \frac{\overline{V}_j}{\overline{V}} = \frac{\mathrm{E}_{\mathbf{Z}}\big[\mathrm{Var}_{U_j}\big[\mathrm{E}_{\mathbf{U}_{(j)}}\big[R(\mathbf{U},\mathbf{Z}) \mid U_j, \mathbf{Z}\big]\big]\big]}{\mathrm{E}_{\mathbf{Z}}\big[\mathrm{Var}_{\mathbf{U}}\big[R(\mathbf{U},\mathbf{Z}) \mid \mathbf{Z}\big]\big]}, \quad j = 1, 2, \ldots, J, \\
\overline{\tau}_j &:= \frac{\overline{V}_j + \sum_{j'=1, j'\neq j}^{J} \overline{V}_{jj'} + \cdots + \overline{V}_{12\cdots J}}{\overline{V}} \\
&= \frac{\mathrm{E}_{\mathbf{Z}}\big[\mathrm{E}_{\mathbf{U}_{(j)}}\big[\mathrm{Var}_{\mathbf{U}_j}\big[R(\mathbf{U},\mathbf{Z}) \mid \mathbf{U}_{(j)}, \mathbf{Z}\big]\big]\big]}{\mathrm{E}_{\mathbf{Z}}\big[\mathrm{Var}_{\mathbf{U}}\big[R(\mathbf{U},\mathbf{Z}) \mid \mathbf{Z}\big]\big]}, \quad j = 1, 2, \ldots, J, \\
\overline{\eta}_j &:= \overline{\tau}_j - \overline{\sigma}_j, \quad j = 1, 2, \ldots, J.
\end{aligned} \tag{5.18}$$

The variance-based sensitivity indices given by Eq. (5.18) are obtained by averaging over the noise factors $\mathbf{Z}$, thus minimizing the effect of experimental variability. We will be referring to these sensitivity indices as *noise-reduced variance-based sensitivity indices* (NR-VSI's). The NR-SESI $\overline{\sigma}_j$ quantifies the fractional singular contribution of the j^{th} biochemical factor to the average response variance $\overline{V}$, the NR-JESI $\overline{\eta}_j$ quantifies the fractional contribution of the j^{th} biochemical factor to $\overline{V}$ jointly with one or more other factors, whereas, the NR-TESI $\overline{\tau}_j$ quantifies the fractional total (singular and joint) contribution of the j^{th} biochemical factor to $\overline{V}$.

By following the previous discussions in Section 3.1, if $\overline{\sigma}_j = \overline{\eta}_j = 0$, then we will conclude that the biochemical factor j does not appreciably influence the system response. On the other hand, if $\overline{\sigma}_j > 0$ and $\overline{\eta}_j = 0$, then we will conclude that the biochemical factor j influences the system response mostly singularly. Moreover, if $\overline{\sigma}_j = 0$ and $\overline{\eta}_j > 0$, we will conclude that the biochemical factor j influences the system response mostly jointly with other biochemical factors, whereas, if $\overline{\sigma}_j > 0$ and $\overline{\eta}_j > 0$, we will conclude that the biochemical factor j influences the system response both singularly and jointly with other biochemical factors. In practice, we set a small threshold θ to determine whether $\overline{\sigma}_j$ and $\overline{\eta}_j$ are sufficiently larger than zero.

5.3 Monte Carlo Estimation

Unfortunately, the NR-VSI's given by Eq. (5.18) cannot be computed analytically. For their evaluation, we must resort to Monte Carlo estimation. By using the variance decomposition formula $\text{Var}_Y\big[Y\big] = \text{Var}_X\big[\text{E}_Y\big[Y|X\big]\big] + \text{E}_X\big[\text{Var}_Y\big[Y|X\big]\big]$, we can show

that

$$\begin{aligned}\overline{\sigma}_j &= \frac{\mathrm{Var}_{U_j,\mathbf{Z}}\big[\mathrm{E}_{\mathbf{U}_{(j)}}\big[R(\mathbf{U},\mathbf{Z}) \mid U_j,\mathbf{Z}\big]\big] - \mathrm{Var}_{\mathbf{Z}}\big[\mathrm{E}_{\mathbf{U}}\big[R(\mathbf{U},\mathbf{Z}) \mid \mathbf{Z}\big]\big]}{\mathrm{Var}_{\mathbf{U},\mathbf{Z}}\big[R(\mathbf{U},\mathbf{Z})\big] - \mathrm{Var}_{\mathbf{Z}}\big[\mathrm{E}_{\mathbf{U}}\big[R(\mathbf{U},\mathbf{Z}) \mid \mathbf{Z}\big]\big]}, \\ \overline{\tau}_j &= \frac{\mathrm{Var}_{\mathbf{U},\mathbf{Z}}\big[R(\mathbf{U},\mathbf{Z})\big] - \mathrm{Var}_{\mathbf{U}_{(j)},\mathbf{Z}}\big[\mathrm{E}_{U_j}\big[R(\mathbf{U},\mathbf{Z}) \mid \mathbf{U}_{(j)},\mathbf{Z}\big]\big]}{\mathrm{Var}_{\mathbf{U},\mathbf{Z}}\big[R(\mathbf{U},\mathbf{Z})\big] - \mathrm{Var}_{\mathbf{Z}}\big[\mathrm{E}_{\mathbf{U}}\big[R(\mathbf{U},\mathbf{Z}) \mid \mathbf{Z}\big]\big]}\,.\end{aligned} \tag{5.19}$$

These formulas allow us to use a Monte Carlo method, similar to the one employed in Section 3.2, based on a Latin hypercube sampling scheme that efficiently samples the random factors in order to reduce estimation variance. We refer to this method as Monte Carlo Latin hypercube sampling (MC-LHS).

The MC-LHS method we use to estimate the NR-SESI's and the NR-TESI's is based on Eq. (5.19) and extends the MC-LHS method used in Section 3.2 by taking into consideration experimental variability. For a given sample size L, the method starts by forming two groups:

$$\begin{array}{|cccccccc|}\hline u_1^{(1)} & u_2^{(1)} & \ldots & u_J^{(1)} & z_1^{(1)} & z_2^{(1)} & \ldots & z_P^{(1)} \\ u_1^{(2)} & u_2^{(2)} & \ldots & u_J^{(2)} & z_1^{(2)} & z_2^{(2)} & \ldots & z_P^{(2)} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ u_1^{(L)} & u_2^{(L)} & \ldots & u_J^{(L)} & z_1^{(L)} & z_2^{(L)} & \ldots & z_P^{(L)} \\ \hline\end{array}$$

$$\begin{array}{|cccccccc|}\hline u_1^{(L+1)} & u_2^{(L+1)} & \ldots & u_J^{(L+1)} & z_1^{(L+1)} & z_2^{(L+1)} & \ldots & z_P^{(L+1)} \\ u_1^{(L+2)} & u_2^{(L+2)} & \ldots & u_J^{(L+2)} & z_1^{(L+2)} & z_2^{(L+2)} & \ldots & z_P^{(L+2)} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ u_1^{(2L)} & u_2^{(2L)} & \ldots & u_J^{(2L)} & z_1^{(2L)} & z_2^{(2L)} & \ldots & z_P^{(2L)} \\ \hline\end{array}$$

of $2L$ Latin hypercube samples of the statistically independent random factors $\mathbf{U} = \big\{U_1, U_2, \ldots, U_J\big\}$ and $\mathbf{Z} = \big\{Z_1, Z_2, \ldots, Z_P\big\}$. The samples are drawn independently from the Gaussian probability densities of $\{U_j, j = 1, 2, \ldots, J\}$ and $\{Z_p, p = 1, 2, \ldots, P\}$.

Subsequently, we group the samples together to form the following values for $\mathbf{U}$ and $\mathbf{Z}$:

$$\mathbf{u}^{(l)} = \left\{u_1^{(l)}, u_2^{(l)}, \ldots, u_J^{(l)}\right\}, \quad l = 1, 2, \ldots, 2L$$

$$\mathbf{u}_j^{(l)} = \left\{u_1^{(L+l)}, \ldots, u_{j-1}^{(L+l)}, u_j^{(l)}, u_{j+1}^{(L+l)}, \ldots, u_J^{(L+l)}\right\}, \quad j = 1, 2, \ldots, J, \quad l = 1, 2, \ldots, L$$

$$\mathbf{u}_{(j)}^{(l)} = \left\{u_1^{(l)}, \ldots, u_{j-1}^{(l)}, u_j^{(L+l)}, u_{j+1}^{(l)}, \ldots, u_J^{(l)}\right\}, \quad j = 1, 2, \ldots, J, \quad l = 1, 2, \ldots, L$$

$$\mathbf{z}^{(l)} = \left\{z_1^{(l)}, z_2^{(l)}, \ldots, z_P^{(l)}\right\}, \quad l = 1, 2, \ldots, 2L\,.$$

We then evaluate the following Monte Carlo estimates for the conditional and unconditional variances in Eq. (5.19):

$$\begin{aligned}\widehat{\mathrm{Var}}_{\mathbf{U},\mathbf{Z}j}\left[R(\mathbf{U},\mathbf{Z})\right] = \frac{1}{8L}\Big[&\sum_{l=1}^{L} R^2(\mathbf{u}^{(l)},\mathbf{z}^{(l)}) + \sum_{l=1}^{L} R^2(\mathbf{u}^{(l)},\mathbf{z}^{(L+l)}) + \sum_{l=1}^{L} R^2(\mathbf{u}^{(L+l)},\mathbf{z}^{(l)}) \\ &+ \sum_{l=1}^{L} R^2(\mathbf{u}^{(L+l)},\mathbf{z}^{(L+l)}) + \sum_{l=1}^{L} R^2(\mathbf{u}_j^{(l)},\mathbf{z}^{(l)}) + \sum_{l=1}^{L} R^2(\mathbf{u}_j^{(l)},\mathbf{z}^{(L+l)}) \\ &+ \sum_{l=1}^{L} R^2(\mathbf{u}_{(j)}^{(l)},\mathbf{z}^{(l)}) + \sum_{l=1}^{L} R^2(\mathbf{u}_{(j)}^{(l)},\mathbf{z}^{(L+l)})\Big] - \widehat{\mathrm{E}}^2_{\mathbf{U},\mathbf{Z}j}\left[R(\mathbf{U},\mathbf{Z})\right],\end{aligned}$$

$$\begin{aligned}\widehat{\mathrm{Var}}_{\mathbf{Z}j}\big[\mathrm{E}_{\mathbf{U}}\big[R(\mathbf{U},\mathbf{Z}) \mid \mathbf{Z}\big]\big] = \frac{1}{4L}\Big[&\sum_{l=1}^{L} R(\mathbf{u}^{(l)},\mathbf{z}^{(l)})R(\mathbf{u}^{(L+l)},\mathbf{z}^{(l)}) \\ &+ \sum_{l=1}^{L} R(\mathbf{u}^{(l)},\mathbf{z}^{(L+l)})R(\mathbf{u}^{(L+l)},\mathbf{z}^{(L+l)}) + \sum_{l=1}^{L} R(\mathbf{u}_j^{(l)},\mathbf{z}^{(l)})R(\mathbf{u}_{(j)}^{(l)},\mathbf{z}^{(l)}) \\ &+ \sum_{l=1}^{L} R(\mathbf{u}_j^{(l)},\mathbf{z}^{(L+l)})R(\mathbf{u}_{(j)}^{(l)},\mathbf{z}^{(L+l)})\Big] - \widehat{\mathrm{E}}^2_{\mathbf{U},\mathbf{Z}j}\left[R(\mathbf{U},\mathbf{Z})\right],\end{aligned}$$

$$\begin{aligned}\widehat{\mathrm{Var}}_{U_j,\mathbf{Z}}\big[\mathrm{E}_{\mathbf{U}_{(j)}}\big[R(\mathbf{U},\mathbf{Z}) \mid U_j,\mathbf{Z}\big]\big] = \frac{1}{4L}\Big[&\sum_{l=1}^{L} R(\mathbf{u}^{(l)},\mathbf{z}^{(l)})R(\mathbf{u}_j^{(l)},\mathbf{z}^{(l)}) \\ &+ \sum_{l=1}^{L} R(\mathbf{u}^{(l)},\mathbf{z}^{(L+l)})R(\mathbf{u}_j^{(l)},\mathbf{z}^{(L+l)}) + \sum_{l=1}^{L} R(\mathbf{u}^{(L+l)},\mathbf{z}^{(l)})R(\mathbf{u}_{(j)}^{(l)},\mathbf{z}^{(l)}) \\ &+ \sum_{l=1}^{L} R(\mathbf{u}^{(L+l)},\mathbf{z}^{(L+l)})R(\mathbf{u}_{(j)}^{(l)},\mathbf{z}^{(L+l)})\Big] - \widehat{\mathrm{E}}^2_{\mathbf{U},\mathbf{Z}j}\left[R(\mathbf{U},\mathbf{Z})\right],\end{aligned}$$

$$\widehat{\mathrm{Var}}_{\mathbf{U}_{(j)},\mathbf{Z}}\big[\mathrm{E}_{U_j}\big[R(\mathbf{U},\mathbf{Z}) \mid \mathbf{U}_{(j)},\mathbf{Z}\big]\big] = \frac{1}{4L}\Big[\sum_{l=1}^{L} R(\mathbf{u}^{(l)},\mathbf{z}^{(l)})R(\mathbf{u}_{(j)}^{(l)},\mathbf{z}^{(l)}) + \sum_{l=1}^{L} R(\mathbf{u}^{(l)},\mathbf{z}^{(L+l)})R(\mathbf{u}_{(j)}^{(l)},\mathbf{z}^{(L+l)}) + \sum_{l=1}^{L} R(\mathbf{u}^{(L+l)},\mathbf{z}^{(l)})R(\mathbf{u}_{j}^{(l)},\mathbf{z}^{(l)}) + \sum_{l=1}^{L} R(\mathbf{u}^{(L+l)},\mathbf{z}^{(L+l)})R(\mathbf{u}_{j}^{(l)},\mathbf{z}^{(L+l)})\Big] - \widehat{\mathrm{E}}^2_{\mathbf{U},\mathbf{Z}j}\left[R(\mathbf{U},\mathbf{Z})\right],$$

where

$$\widehat{\mathrm{E}}^2_{\mathbf{U},\mathbf{Z}j}\left[R(\mathbf{U},\mathbf{Z})\right] = \frac{1}{4L}\Big[\sum_{l=1}^{L} R(\mathbf{u}^{(l)},\mathbf{z}^{(l)})R(\mathbf{u}^{(L+l)},\mathbf{z}^{(L+l)}) + \sum_{l=1}^{L} R(\mathbf{u}^{(l)},\mathbf{z}^{(L+l)})R(\mathbf{u}^{(L+l)},\mathbf{z}^{(l)}) + \sum_{l=1}^{L} R(\mathbf{u}_{j}^{(l)},\mathbf{z}^{(l)})R(\mathbf{u}_{(j)}^{(l)},\mathbf{z}^{(L+l)}) + \sum_{l=1}^{L} R(\mathbf{u}_{j}^{(l)},\mathbf{z}^{(L+l)})R(\mathbf{u}_{(j)}^{(l)},\mathbf{z}^{(l)})\Big].$$

Finally, we estimate the NR-VSI's by:

$$\widehat{\overline{\sigma}}_j \simeq \frac{\widehat{\mathrm{Var}}_{U_j,\mathbf{Z}}\big[\mathrm{E}_{\mathbf{U}_{(j)}}\big[R(\mathbf{U},\mathbf{Z}) \mid U_j,\mathbf{Z}\big]\big] - \widehat{\mathrm{Var}}_{\mathbf{Z}j}\big[\mathrm{E}_{\mathbf{U}}\big[R(\mathbf{U},\mathbf{Z}) \mid \mathbf{Z}\big]\big]}{\widehat{\mathrm{Var}}_{\mathbf{U},\mathbf{Z}j}\big[R(\mathbf{U},\mathbf{Z})\big] - \widehat{\mathrm{Var}}_{\mathbf{Z}j}\big[\mathrm{E}_{\mathbf{U}}\big[R(\mathbf{U},\mathbf{Z}) \mid \mathbf{Z}\big]\big]},$$

$$\widehat{\overline{\tau}}_j \simeq \frac{\widehat{\mathrm{Var}}_{\mathbf{U},\mathbf{Z}j}\big[R(\mathbf{U},\mathbf{Z})\big] - \widehat{\mathrm{Var}}_{\mathbf{U}_{(j)},\mathbf{Z}}\big[\mathrm{E}_{U_j}\big[R(\mathbf{U},\mathbf{Z}) \mid \mathbf{U}_{(j)},\mathbf{Z}\big]\big]}{\widehat{\mathrm{Var}}_{\mathbf{U},\mathbf{Z}j}\big[R(\mathbf{U},\mathbf{Z})\big] - \widehat{\mathrm{Var}}_{\mathbf{Z}j}\big[\mathrm{E}_{\mathbf{U}}\big[R(\mathbf{U},\mathbf{Z}) \mid \mathbf{Z}\big]\big]}.$$

Note that computation of the previous estimators requires evaluation of the system responses $R(\mathbf{u}^{(l)},\mathbf{z}^{(l)})$, $R(\mathbf{u}^{(l)},\mathbf{z}^{(L+l)})$, $R(\mathbf{u}^{(L+l)},\mathbf{z}^{(l)})$, and $R(\mathbf{u}^{(L+l)},\mathbf{z}^{(L+l)})$, for $l = 1, 2, \ldots, L$, as well as $R(\mathbf{u}_j^{(l)},\mathbf{z}^{(l)})$, $R(\mathbf{u}_j^{(l)},\mathbf{z}^{(L+l)})$, $R(\mathbf{u}_{(j)}^{(l)},\mathbf{z}^{(l)})$, and $R(\mathbf{u}_{(j)}^{(l)},\mathbf{z}^{(L+l)})$, for $j = 1, 2, \ldots, J$, $l = 1, 2, \ldots, L$. This corresponds to a total of $4L + 4LJ = 4L(J+1)$ system evaluations, by solving the system of differential equations given by Eq. (2.2), where L is the number of Latin hypercube samples used and J is the number of biochemical factors considered ($J = M$ for ROSA and $J = N$ for SOSA). It is worthwhile noticing that to estimate the variance-based sensitivity indices given in Section 3.1, which do not consider experimental variability, we need $2L(J+1)$ system evaluations

by the Monte Carlo method given in Section 3.2. If the number of Latin hypercube samples L is the same in both cases, then the computational time required for estimating the NR-VSI's is only twice as much as the time required for estimating the sensitivity indices without experimental variability. On the other hand, if we directly use Eq. (5.18) to estimate the NR-VSI's by averaging the conditional variances obtained at different sampled values of the noise factors $\mathbf{Z}$, then the total number of required system evaluations will be $2LL_z(J+1)$, where L_z is the number of $\mathbf{Z}$ samples used. This will be appreciably larger than $4L(J+1)$ (since $L_z \gg 2$) and prohibitively expensive for large biochemical reaction systems.

5.4 Derivative Approximation

If, for every $\mathbf{z}$, the response function $R(\mathbf{u}, \mathbf{z})$ is sufficiently smooth around $(\mathbf{0}, \mathbf{z})$, so that its derivatives of orders ≥ 3 at $(\mathbf{0}, \mathbf{z})$ are negligible, and if the biochemical factors $\mathbf{U}$ are i.i.d. zero-mean Gaussian random variables with sufficiently small standard deviations λ, such that

$$\lambda^4 \left[\frac{\partial^2 R(\mathbf{0}, \mathbf{z})}{\partial u_j \partial u_{j'}} \right]^2 \simeq 0, \quad \text{for every} \;\; j, j' = 1, 2, \ldots, J, \tag{5.20}$$

then, by following a similar procedure as the one we used to derive Eq. (3.10), we have

$$V(\mathbf{z}) = \mathrm{Var}_{\mathbf{U}}\big[R(\mathbf{U}, \mathbf{z})\big] \simeq \sum_{j=1}^{J} V_j(\mathbf{z}), \tag{5.21}$$

where

$$V_j(\mathbf{z}) = \mathrm{Var}_{U_j}\big[\mathrm{E}_{\mathbf{U}_{(j)}}\big[R(\mathbf{U}, \mathbf{z}) \mid U_j\big]\big] \simeq \lambda^2 \left[\frac{R(\mathbf{0}, \mathbf{z})}{\partial u_j} \right]^2. \tag{5.22}$$

In this case,

$$\overline{V} \simeq \overline{V}_1 + \overline{V}_1 + \cdots + \overline{V}_j, \tag{5.23}$$

with

$$\overline{V}_j \simeq \lambda^2 \, \mathrm{E}_{\mathbf{Z}}\Big[\big[\partial R(\mathbf{0}, \mathbf{Z})/\partial u_j\big]^2\Big], \tag{5.24}$$

and the NR-SESI's and NR-TESI's will satisfy

$$\overline{\sigma}_j \simeq \overline{\tau}_j \simeq \frac{\mathrm{E}_{\mathbf{Z}}\big[\big[\partial R(\mathbf{0}, \mathbf{Z})/\partial u_j\big]^2\big]}{\sum_{j'=1}^{J} \mathrm{E}_{\mathbf{Z}}\big[\big[\partial R(\mathbf{0}, \mathbf{Z})/\partial u_{j'}\big]^2\big]}, \quad j = 1, 2, \ldots, J. \tag{5.25}$$

Eq. (5.25) shows that, under certain conditions, the two noise-reduced sensitivity indices considered in this chapter can be approximated by using the means, with respect to the noise factors $\mathbf{Z}$, of the squares of the response derivatives $\partial R(\mathbf{0}, \mathbf{Z})/\partial u_j$, $j = 1, 2, \ldots, J$. The resulting approximation is related to the derivative-based global sensitivity measures recently suggested in the literature by Kucherenko *et al.* [106]. Therefore, we can view the sensitivity analysis approach suggested by Kucherenko *et al.* as a direct consequence of the approach discussed in this chapter. However, the sensitivity analysis results obtained by the Kucherenko approach will be legitimate only if the underlying assumptions that lead to Eq. (5.25) are satisfied, which is not easy to verify in practice. Note also that Eq. (5.25) implies that sensitivity analysis based on the derivative approximation given by Eq. (5.25) is limited to the case when the biochemical factors contribute to the response variance mostly singularly, with all joint contributions being negligible. This limits the scope of a sensitivity analysis approach based on the approximation given by Eq. (5.25), unless the response function is nearly additive with respect to the biochemical factors.[18] Finally, to evaluate the derivative-

[18] If the system response function $R(\mathbf{u}, \mathbf{z})$ is additive with respect to $\mathbf{u}$, then we can show that Eq. (5.23) is exact.

based approximation given by Eq. (5.25) we must also resort to Monte Carlo estimation. This requires $2L_zJ$ system evaluations (provided that a central finite-difference approximation of the derivative is used), as compared to the $4L(J+1)$ system evaluations required to estimate the NR-VSI's by Eq. (5.19), which will be beneficial only if $L_z \ll L$.

5.5 Dimensionality Reduction

A serious bottleneck when calculating sensitivity indices by Monte Carlo estimation is the substantial computational time required to achieve a certain level of accuracy. This is due to the fact that accurate Monte Carlo estimation requires a large number of system evaluations, which is done by integrating the differential equations given by Eq. (2.2) at each Monte Carlo step. To make matters worse, stiffness may substantially increase the computational time required for integrating these equations. Since estimation of the NR-VSI's given by Eq. (5.19) requires $4L(J+1)$ system evaluations, where J is the number of biochemical factors used for sensitivity analysis, we may be able to appreciably reduce computational time by considering a smaller number of biochemical factors in our analysis, thus reducing the dimensionality of the problem at hand. One way to accomplish this goal is to use a relatively fast method to first identify non-influential biochemical factors, and subsequently employ a variance-based sensitivity analysis method that focuses on the influential factors, while treating all non-influential biochemical factors as a single group [23].

Variance-based sensitivity analysis can be easily tailored to accommodate groups

of factors [79, 102]. Grouping factors allows us to reduce computations at the cost of losing information about the relative importance of factors within a group. Since we are not interested in the relative importance of non-influential biochemical factors, we can reduce the computational burden of variance-based sensitivity analysis by treating these factors as a group. Considering also the fact that, in most biochemical reaction systems, only a small number of biochemical factors turn out to be influential, we may achieve substantial computational savings by aggregating the non-influential biochemical factors into one group.

Let us denote by $\mathbf{U}_0$ and $\mathbf{U}_1$ the non-influential and influential biochemical factors, respectively. Note that $\mathbf{U} = \{\mathbf{U}_0, \mathbf{U}_1\}$. Moreover, let J_0 be the total number of non-influential biochemical factors, in which case, the total number of influential biochemical factors will be $J_1 = J - J_0$. Although the NR-VSI's associated with the influential biochemical factors will still be given by Eq. (5.18), or Eq. (5.19), the NR-VSI's associated with the group of non-influential factors will now be given by

$$\begin{aligned}
\overline{\sigma}_0 &= \frac{\mathrm{E}_{\mathbf{Z}}\big[\mathrm{Var}_{\mathbf{U}_0}\big[\mathrm{E}_{\mathbf{U}_1}\big[R(\mathbf{U}_0, \mathbf{U}_1, \mathbf{Z}) \mid \mathbf{U}_0, \mathbf{Z}\big]\big]\big]}{\mathrm{E}_{\mathbf{Z}}\big[\mathrm{Var}_{\mathbf{U}_0,\mathbf{U}_1}\big[R(\mathbf{U}_0, \mathbf{U}_1, \mathbf{Z})\big]\big]} \\
&= \frac{\mathrm{Var}_{\mathbf{U}_0,\mathbf{Z}}\big[\mathrm{E}_{\mathbf{U}_1}\big[R(\mathbf{U}_0, \mathbf{U}_1, \mathbf{Z}) \mid \mathbf{U}_0, \mathbf{Z}\big]\big] - \mathrm{Var}_{\mathbf{Z}}\big[\mathrm{E}_{\mathbf{U}_0,\mathbf{U}_1}\big[R(\mathbf{U}_0, \mathbf{U}_1, \mathbf{Z}) \mid \mathbf{Z}\big]\big]}{\mathrm{Var}_{\mathbf{U}_0,\mathbf{U}_1,\mathbf{Z}}\big[R(\mathbf{U}_0, \mathbf{U}_1, \mathbf{Z})\big] - \mathrm{Var}_{\mathbf{Z}}\big[\mathrm{E}_{\mathbf{U}_0,\mathbf{U}_1}\big[R(\mathbf{U}_0, \mathbf{U}_1, \mathbf{Z}) \mid \mathbf{Z}\big]\big]}, \\
\overline{\tau}_0 &= \frac{\mathrm{E}_{\mathbf{Z}}\big[\mathrm{E}_{\mathbf{U}_1}\big[\mathrm{Var}_{\mathbf{U}_0}\big[R(\mathbf{U}_0, \mathbf{U}_1, \mathbf{Z}) \mid \mathbf{U}_1, \mathbf{Z}\big]\big]\big]}{\mathrm{E}_{\mathbf{Z}}\big[\mathrm{Var}_{\mathbf{U}_0,\mathbf{U}_1}\big[R(\mathbf{U}_0, \mathbf{U}_1, \mathbf{Z})\big]\big]} \\
&= \frac{\mathrm{Var}_{\mathbf{U}_0,\mathbf{U}_1,\mathbf{Z}}\big[R(\mathbf{U}_0, \mathbf{U}_1, \mathbf{Z})\big] - \mathrm{Var}_{\mathbf{U}_1,\mathbf{Z}}\big[\mathrm{E}_{\mathbf{U}_0}\big[R(\mathbf{U}_0, \mathbf{U}_1, \mathbf{Z}) \mid \mathbf{U}_1, \mathbf{Z}\big]\big]}{\mathrm{Var}_{\mathbf{U}_0,\mathbf{U}_1,\mathbf{Z}}\big[R(\mathbf{U}_0, \mathbf{U}_1, \mathbf{Z})\big] - \mathrm{Var}_{\mathbf{Z}}\big[\mathrm{E}_{\mathbf{U}_0,\mathbf{U}_1}\big[R(\mathbf{U}_0, \mathbf{U}_1, \mathbf{Z}) \mid \mathbf{Z}\big]\big]}, \\
\overline{\eta}_0 &= \overline{\tau}_0 - \overline{\sigma}_0.
\end{aligned} \tag{5.26}$$

The MC-LHS estimators given in the Section 5.3 can be easily modified to estimate the NR-VSI's $\overline{\sigma}_0$ and $\overline{\tau}_0$ given by Eq. (5.26). It turns out that, in this case, computation of

all NR-VSI's requires $4L(J-J_0+2)=4L(J_1+2)$ system evaluations, which is smaller than the $4L(J_0+J_1+1)$ system evaluations required when the non-influential factors are not grouped together. As a consequence, grouping the non-influential factors results in a reduction of system evaluations by a factor of $1+(J_0-1)/(J_1+2)$, which can be substantial when the number of influential factors J_1 is much smaller than the number of non-influential factors J_0.

To employ the previous idea, we must develop a method that we can use to efficiently identify non-influential biochemical factors before we perform variance-based sensitivity analysis using Monte Carlo estimation. Let us define the *second-order* NR-VSI's

$$\begin{aligned}
\overline{\sigma}_j^{(2)} &:= \frac{\overline{V}_j}{\sum_{j=1}^{J}\overline{V}_j+\sum_{j=1}^{J-1}\sum_{j'=j+1}^{J}\overline{V}_{jj'}},\\
\overline{\tau}_j^{(2)} &:= \frac{\overline{V}_j+\sum_{j'=1,j'\neq j}^{J}\overline{V}_{jj'}}{\sum_{j=1}^{J}\overline{V}_j+\sum_{j=1}^{J-1}\sum_{j'=j+1}^{J}\overline{V}_{jj'}},\\
\overline{\eta}_j^{(2)} &:= \overline{\tau}_j^{(2)}-\overline{\sigma}_j^{(2)}=\frac{\sum_{j'=1,j'\neq j}^{J}\overline{V}_{jj'}}{\sum_{j=1}^{J}\overline{V}_j+\sum_{j=1}^{J-1}\sum_{j'=j+1}^{J}\overline{V}_{jj'}},
\end{aligned} \tag{5.27}$$

and the *second-order* group NR-VSI's

$$\begin{aligned}
\overline{\sigma}_0^{(2)} &:= \frac{\sum_{j\in\mathbf{U}_0}\overline{V}_j}{\sum_{j=1}^{J}\overline{V}_j+\sum_{j=1}^{J-1}\sum_{j'=j+1}^{J}\overline{V}_{jj'}},\\
\overline{\tau}_0^{(2)} &:= \frac{\sum_{\{j:U_j\in\mathbf{U}_0\}}\overline{V}_j+\sum\sum_{1\leq j<j'\leq J,\{j:U_j\in\mathbf{U}_0\}\text{ or }\{j':U_{j'}\in\mathbf{U}_0\}}\overline{V}_{jj'}}{\sum_{j=1}^{J}\overline{V}_j+\sum_{j=1}^{J-1}\sum_{j'=j+1}^{J}\overline{V}_{jj'}},\\
\overline{\eta}_0^{(2)} &:= \overline{\tau}_0^{(2)}-\overline{\sigma}_0^{(2)}=\frac{\sum\sum_{1\leq j<j'\leq J,\{j:U_j\in\mathbf{U}_0\}\text{ or }\{j':U_{j'}\in\mathbf{U}_0\}}\overline{V}_{jj'}}{\sum_{j=1}^{J}\overline{V}_j+\sum_{j=1}^{J-1}\sum_{j'=j+1}^{J}\overline{V}_{jj'}},
\end{aligned} \tag{5.28}$$

obtained from the corresponding NR-VSI's and group NR-VSI's, given by Eq. (5.18) and Eq. (5.26), by setting all third-order and higher-order variance terms equal to zero. It has been argued in the literature that, in most well-defined physical systems, appreciable high-order interactions normally occur only among those biochemical factors that already demonstrate substantial low-order interactions [108]. We can capitalize

on this argument and assume that a biochemical factor is non-influential if and only if its singular and joint, with another factor, contribution to the response variance is negligible. As a consequence, we may say that the biochemical factor U_j is non-influential if and only if $\overline{\tau}_j^{(2)} \simeq 0$. Moreover, if $\overline{\tau}_0^{(2)} \simeq 0$ [or $\overline{\sigma}_0^{(2)} \simeq 0$ and $\overline{\eta}_0^{(2)} \simeq 0$], then we may conclude that all biochemical factors contained in $\mathbf{U}_0$ are indeed non-influential.

Although a number of techniques can be adopted for estimating the second-order NR-VSI's and second-order group NR-VSI's, we will consider here the Orthonormal Hermite Approximation (OHA) discussed in Chapter 4, based on approximating the response function of a biochemical reaction system by orthonormal Hermite polynomials. It has been demonstrated in Chapter 4 (also in [93]) that this method can provide good approximations of second-order variance-based sensitivity indices. If we set $\mathbf{W} = \{\mathbf{U}, \mathbf{Z}\}$, then OHA employs the following approximation of the system response function $R(\mathbf{W}) = R(\mathbf{U}, \mathbf{Z})$:

$$\begin{aligned}
\widehat{R}(\mathbf{w}) = \widehat{\rho}_0 \;\; &+ \sum_{q=1}^{J+P} \left[\alpha_{q,1} \frac{w_q}{s_q} + \frac{\alpha_{q,2}}{\sqrt{2}} \left(\frac{w_q^2}{s_q^2} - 1 \right) \right] \\
&+ \sum_{q=1}^{J+P-1} \sum_{q'=q+1}^{J+P} \alpha_{qq',1} \frac{w_q w_{q'}}{s_q s_{q'}} \\
&+ \sum_{q=1}^{J+P-1} \sum_{q'=q+1}^{J+P} \frac{\alpha_{qq',2}}{\sqrt{2}} \left(\frac{w_q^2}{s_q^2} - 1 \right) \frac{w_{q'}}{s_{q'}} \\
&+ \sum_{q=1}^{J+P-1} \sum_{q'=q+1}^{J+P} \frac{\alpha_{qq',3}}{\sqrt{2}} \frac{w_q}{s_q} \left(\frac{w_{q'}^2}{s_{q'}^2} - 1 \right) \\
&+ \sum_{q=1}^{J+P-1} \sum_{q'=q+1}^{J+P} \frac{\alpha_{qq',4}}{2} \left(\frac{w_q^2}{s_q^2} - 1 \right) \left(\frac{w_{q'}^2}{s_{q'}^2} - 1 \right),
\end{aligned} \tag{5.29}$$

where $w_q = u_q$, for $q = 1, 2, \ldots, J$, whereas, $w_q = z_{q-J}$, for $q = J+1, J+2, \ldots, J+P$. Moreover, $s_q = \lambda^{\ddagger}$, for $q = 1, 2, \ldots, J$, and $s_q = \sigma$, for $q = J+1, J+2, \ldots, J+P$, in the case of ROSA, whereas, $s_q = \lambda$, for $q = 1, 2, \ldots, J$, and $s_q = \sigma$, for $q = J+1, J+$

$2, \ldots, J+P$, in the case of SOSA. Note that $\widehat{R}(\mathbf{w})$ provides a polynomial approximation of the response function $R(\mathbf{w})$ in terms of first- and second-order orthonormal Hermite polynomials. In Eq. (5.29), the values of the parameters $\widehat{\rho}_0$ and α must be appropriately determined so that $\widehat{R}(\mathbf{w})$ sufficiently approximates the response function $R(\mathbf{w})$ over the entire space of biochemical and experimental factor values. This can be done by polynomial regression based on the Monte Carlo Latin hypercube sampling discussed in Section 4.5.4.

By using Eq. (5.29), the orthonormality of the Hermite polynomials, the statistical independence and zero-mean Gaussianity of $\mathbf{U}$ and $\mathbf{Z}$, we can show that the second-order NR-VSI's, given by Eq. (5.27), can be approximated by

$$\begin{aligned}
\widehat{\overline{\sigma}}_j^{(2)} &= \frac{\overline{V}_j}{\sum_{j=1}^{J} \overline{V}_j + \sum_{j=1}^{J-1} \sum_{j'=j+1}^{J} \overline{V}_{jj'}}, \\
\widehat{\overline{\tau}}_j^{(2)} &= \frac{\overline{V}_j + \sum_{j'=1, j' \neq j}^{J} \overline{V}_{jj'}}{\sum_{j=1}^{J} \overline{V}_j + \sum_{j=1}^{J-1} \sum_{j'=j+1}^{J} \overline{V}_{jj'}}, \\
\widehat{\overline{\eta}}_j^{(2)} &= \frac{\sum_{j'=1, j' \neq j}^{J} \overline{V}_{jj'}}{\sum_{j=1}^{J} \overline{V}_j + \sum_{j=1}^{J-1} \sum_{j'=j+1}^{J} \overline{V}_{jj'}},
\end{aligned} \tag{5.30}$$

and the second-order group NR-VSI's, given by Eq. (5.28), can be approximated by

$$\begin{aligned}
\widehat{\overline{\sigma}}_0^{(2)} &= \frac{\sum_{j \in \mathbf{U}_0} \widehat{\overline{V}}_j}{\sum_{j=1}^{J} \widehat{\overline{V}}_j + \sum_{j=1}^{J-1} \sum_{j'=j+1}^{J} \widehat{\overline{V}}_{jj'}}, \\
\widehat{\overline{\tau}}_0^{(2)} &= \frac{\sum_{\{j: U_j \in \mathbf{U}_0\}} \widehat{\overline{V}}_j + \sum \sum_{1 \leq j < j' \leq J, \{j: U_j \in \mathbf{U}_0\} \text{ or } \{j': U_{j'} \in \mathbf{U}_0\}} \widehat{\overline{V}}_{jj'}}{\sum_{j=1}^{J} \widehat{\overline{V}}_j + \sum_{j=1}^{J-1} \sum_{j'=j+1}^{J} \widehat{\overline{V}}_{jj'}}, \\
\widehat{\overline{\eta}}_0^{(2)} &= \frac{\sum \sum_{1 \leq j < j' \leq J, \{j: U_j \in \mathbf{U}_0\} \text{ or } \{j': U_{j'} \in \mathbf{U}_0\}} \widehat{\overline{V}}_{jj'}}{\sum_{j=1}^{J} \widehat{\overline{V}}_j + \sum_{j=1}^{J-1} \sum_{j'=j+1}^{J} \widehat{\overline{V}}_{jj'}},
\end{aligned} \tag{5.31}$$

where

$$\begin{aligned}
\widehat{\overline{V}}_j &= \alpha_{j,1}^2 + \alpha_{j,2}^2 + \sum_{q=J+1}^{J+P} \left(\alpha_{jq,1}^2 + \alpha_{jq,2}^2 + \alpha_{jq,3}^2 + \alpha_{jq,4}^2\right) \\
\widehat{\overline{V}}_{jj'} &= \alpha_{j'j,1}^2 + \alpha_{j'j,2}^2 + \alpha_{j'j,3}^2 + \alpha_{j'j,4}^2, \quad \text{for} \quad j' < j \\
\widehat{\overline{V}}_{jj'} &= \alpha_{jj',1}^2 + \alpha_{jj',2}^2 + \alpha_{jj',3}^2 + \alpha_{jj',4}^2, \text{for} \quad j' > j \, . \qquad (5.32)
\end{aligned}$$

After evaluating these approximations, we can use them to rank all biochemical factors U_j in accordance to the estimated $\widehat{\overline{\tau}}_j^{(2)}$ values and cluster the J_0 biochemical factors with the smallest $\widehat{\overline{\tau}}_j^{(2)}$ values in a group $\mathbf{U}_0$, such that the corresponding values of $\widehat{\overline{\sigma}}_0^{(2)}$ and $\widehat{\overline{\eta}}_0^{(2)}$ are both below a predefined small threshold ϵ. To maximize dimensionality reduction, we must find the largest possible number J_0 of biochemical factors that we can place in $\mathbf{U}_0$. Finally, we can employ MC-LHS and Eqs. (5.19) and (5.26), to estimate the NR-VSI's of the influential factors $\mathbf{U}_1 = \mathbf{U} \setminus \mathbf{U}_0$, as well as the group NR-VSI's of the non-influential factors.

The previous approach depends on the validity of our assumption that a biochemical factor is influential only if it appreciably affects the system response at least singularly or jointly with another factor. If this assumption is violated, then the classification of the biochemical factors obtained by OHA may not be accurate. However, we can numerically validate the accuracy of classification by checking the estimated value of the group NR-SESI $\overline{\sigma}_0$ and group NR-JESI $\overline{\eta}_0$. If $\overline{\sigma}_0$ and $\overline{\eta}_0$ are both sufficiently small, then the biochemical factors in $\mathbf{U}_0$, determined by OHA, will have (as a group) a negligible contribution to the system response and, thus, will indeed be non-influential. If, however, the estimated value of $\overline{\sigma}_0$ or $\overline{\eta}_0$ is large, then some factors in $\mathbf{U}_0$ will be influential. In this case, we must un-group the factors in $\mathbf{U}_0$ and repeat our grouping

method by appropriately reducing J_0. We expect that, for most biochemical reaction systems of interest to systems biology, the chance for this to happen will be small.

Estimating the second-order NR-VSI's and group NR-VSI's by OHA requires L_0 system evaluations, where L_0 is the number of regression points obtained by Monte Carlo Latin hypercube sampling. L_0 must be no less than the number $2(J+P)^2+1$ of coefficients in Eq. (5.29). Therefore, L_0 increases with the total number of biochemical and experimental factors. Normally, by setting L_0 to be $2 \sim 4$ times the number of unknown coefficients is adequate to obtain a sufficiently accurate fit of the response function, although this clearly depends on the particular response function at hand. In this chapter, we set $L_0 = 8(J+P)^2$. As a consequence, the total number of system evaluations when using OHA-based dimensionality reduction is $4L(J_1+2)+8(J_0+J_1+P)^2$, where J_0, J_1 are the numbers of non-influential and influential biochemical factors, respectively, and P is the number of noise factors. This can be significantly smaller than the number $4L(J_0+J_1+1)$ of system evaluations required in the absence of dimensionality reduction, provided that $(J_0+J_1+P)^2 \ll L(J_0-1)/2$, which is usually true since L is an appreciably large number.

5.6 Numerical Results

We now employ the MAPK signaling cascade model we described in Section 2.2 to demonstrate the noise-reduced variance-based sensitivity analysis technique proposed in this chapter. The sensitivity analysis approach we consider here is based on quantifying the influence of a reaction (for ROSA) or molecular species (for SOSA) on the

duration, integrated response, and strength of ERK-PP activity, which are estimated based on the dynamic behavior of ERK-PP within a time frame of 6 hours. We investigate the sensitivity properties of the MAPK signaling cascade by classifying reactions and molecular species into one of four categories of interest: singularly influential, jointly influential, singularly/jointly influential, and non-influential. We do this by comparing their NR-SESI and NR-JESI values to a threshold $\theta = 0.1$. In the MC-LHS estimation of variance-based sensitivity indices, we set $L = 6{,}000$.

To investigate the effect of different nominal reaction rate values on the results obtained by the variance-based and noise-reduced variance-based sensitivity analysis approaches discussed in this dissertation, we first perturb the published nominal rate constant values (see Tables 2.1 and 2.2) by using Eq. (5.6), where $\mathbf{Z}$ is i.i.d. Gaussian experimental noise with standard deviation $\sigma = 0.4$. We randomly select a set of perturbed rate constant values as the new nominal values, perform variance-based and noise-reduced variance-based sensitivity analysis, and compare the results with the ones obtained by using the published nominal values. To save space, we only present the results obtained for the duration of ERK-PP. We have tested other biochemical reaction systems and other response functions and obtained similar results as the ones presented here; e.g., see [111].

In the first row of Fig. 5.1, we depict the SESI and JESI values for the duration of ERK-PP activity estimated by ROSA with standard deviation $\lambda^{\ddagger} = 0.1$, based on the published (circles) and perturbed (squares) nominal rate values. When using the published rate values, the duration of ERK-PP activity is primarily influenced by reactions 4, 6, and 13 (refer to Fig. 2.1, Table 2.1, and Table 2.2 for identifying these

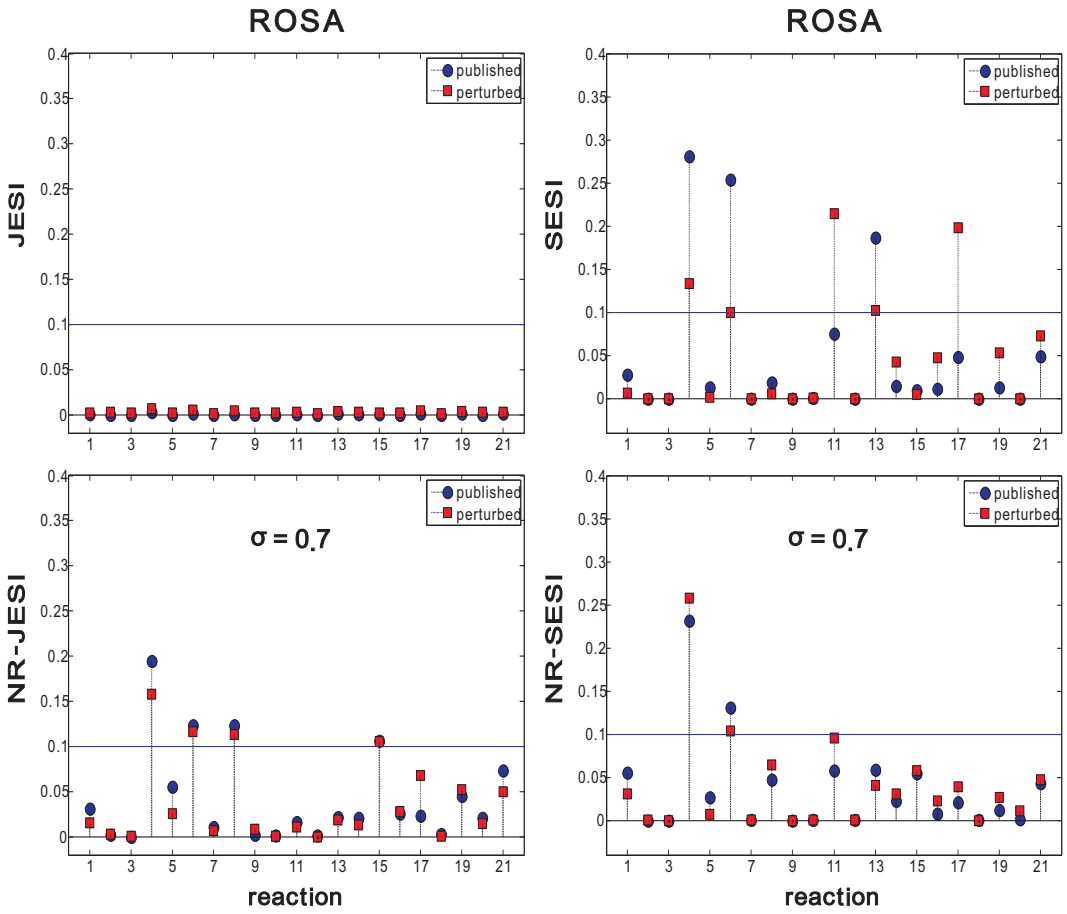

Figure 5.1: ROSA results for the duration of ERK-PP activity in the MAPK signaling cascade based on the published and perturbed nominal rate values with $\lambda^{\ddagger} = 0.1$.

reactions), which exert their influence only singularly (since their SESI values are above the threshold $\theta = 0.1$, whereas the corresponding JESI values are below 0.1). When using the perturbed rate values, the importance of reactions 4, 6, and 13 diminishes significantly, with the SESI values associated with reactions 6 and 13 being above the threshold only marginally. On the other hand, reactions 11 and 17 become the two most influential reactions, since their SESI values surpass the SESI value associated with reaction 4 (which is the most influential reaction under the published rate values). Moreover, reactions 11 and 17 become the most influential ones among all other reactions. Note that reactions 11 and 17 are identified as being non-influential under

the published rate values, which is in sharp contrast with the conclusion made when using the perturbed rate values.

In the first row of Fig. 5.2, found in the Appendix at the end of this chapter, we depict similar results obtained with SOSA. It is therefore quite clear that inconsistencies in sensitivity analysis results, due to experimental variability in the nominal values of the rate constants, can greatly reduce the applicability of variance-based sensitivity analysis.

We can address the previous problem by employing the noise reduction technique we have introduced in this chapter. We illustrate this in the second row of Fig. 5.1 (as well as in the second row of Fig. 5.2), in which we depict the NR-JESI and NR-SESI values associated with the duration of ERK-PP activity estimated by ROSA with $\lambda^{\ddagger} = 0.1$ and $\sigma = 0.7$, based on the published (circles) and perturbed (squares) nominal rate values. Note that, in this case, the values of the two sensitivity indices obtained by using the published and perturbed rate constant values as the nominal values are very close to each other. As a matter of fact, the largest difference is only 4%, which occurs for the NR-JESI of reaction 17, a reaction that is classified as being non-influential in both cases. The results depicted in the second row of Fig. 5.1 indicate that, for both the published and perturbed rate values, the duration of ERK-PP is primarily affected by reactions 4, 6, 8, and 15, with reactions 4 and 6 exerting their influence both singularly and jointly, and with reactions 8 and 15 influencing the duration only jointly. As a consequence, it is clear that the NR-VSI's results are less sensitive to the particular choice of the nominal values and can provide much more consistent results than the traditional variance-based sensitivity indices.

Table 5.1: Number of required system evaluations and equations used for each noise-reduced variance-based sensitivity analysis method considered in this chapter.

method	**required system evaluations**	**equations used**
MC-LHS	$4L(J+1) = 4L(J_0 + J_1 + 1)$	Eq. (5.19)
NR-DA	$2L_z J = 2L_z(J_0 + J_1)$	Eq. (5.25)
OHA-DR	$4L(J_1 + 2) + 8(J_0 + J_1 + P)^2$	Eqs. (5.19), (5.26), (5.30)–(5.32)

L: number of Monte Carlo Latin hypercube samples in MC-LHS
L_z: number of Monte Carlo Latin hypercube samples in NR-DA
J: number of biochemical factors
J_0: number of non-influential biochemical factors in OHA-DR
J_1: number of influential biochemical factors in OHA-DR
P: number of noise factors

In the remaining of this section, we compare the three methods for estimating the NR-VSI's discussed in this chapter, namely, the Monte Carlo Latin hypercube sampling (MC-LHS) method discussed in Section 5.3, the noise-reduced derivative approximation (NR-DA) method discussed in Section 5.4, and the OHA-based dimensionality reduction (OHA-DR) method discussed in Section 5.5. In Table 5.1, we summarize the number of system evaluations and the equations used by each method. For the NR-DA method we employ $L_z = 1{,}000$ Latin hypercube samples to estimate the expectations required by Eq. (5.25). For both ROSA ($J = 21$) and SOSA ($J = 23$), the number of system evaluations required by the derivative approximation method is about 8% of that required by MC-LHS. When using OHA-DR, we employ a threshold $\epsilon = 0.1$ for identifying the non-influential group $\mathbf{U}_0$. The number of system evaluations required by OHA-DR depends on the values of J_0 and J_1, which may vary for different system responses.

In Tables 5.2 and 5.3, we summarize the ROSA-based NR-SESI and NR-JESI values for the duration, integrated response, and strength of ERK-PP activity, estimated by the three methods considered in this chapter (i.e., MC-LHS, NR-DA, and OHA-DR), with standard deviations $\lambda^{\ddagger} = 0.1$, $\sigma = 0.1$, in Table 5.2, and $\lambda^{\ddagger} = 0.4$, $\sigma = 0.7$, in Table 5.3, based on the published nominal values. Moreover, we summarize the corresponding SOSA-based results in Tables 5.6 and 5.7 of the Appendix at the end of this chapter. The results are given in percentages and have been truncated to the nearest integers. To reduce the size of the tables, we depict only the results associated with reactions or molecular species whose truncated NR-SESI or NR-JESI values, estimated by MC-LHS, are at least 0.05. We consider the NR-SESI and NR-JESI values obtained by MC-LHS as the "true" values. Bold reaction or species numbers indicate NR-SESI or NR-JESI values, obtained by MC-LHS, that are larger than 0.1, in which case, the corresponding reactions or molecular species are deemed to be influential.

For the duration, when $\lambda^{\ddagger} = 0.1$ and $\sigma = 0.1$, ROSA-based OHA-DR leads to $J_0 = 13$ and $J_1 = 8$, which means that the number of required system evaluations to estimate the NR-VSI's is about 51% of that required by MC-LHS. For the integrated response and strength, the number of system evaluations required to estimate the NR-VSI's is about 37% of that required by MC-LHS. On the other hand, when $\lambda^{\ddagger} = 0.1$ and $\sigma = 0.1$, the number of system evaluations required by SOSA-based OHA-DR to estimate the NR-VSI's for the duration, integrated response, and strength, is about 37%, 30%, and 30% of that required by MC-LHS, respectively. Inspection of the estimated group values $\overline{\sigma}_0$ and $\overline{\eta}_0$ reveals that the biochemical factor group $\mathbf{U}_0$ determined by OHA is truly non-influential, since both $\overline{\sigma}_0$ and $\overline{\eta}_0$ are less than 0.1 in all cases.

Table 5.2: ROSA-based NR-VSI's for the *duration*, *integrated response*, and *strength* of ERK-PP activity in the MAPK signaling cascade, estimated by the three methods considered in this chapter, with $\lambda^{\ddagger} = 0.1$ and $\sigma = 0.1$. Bold reaction numbers indicate NR-SESI or NR-JESI values, obtained by MC-LHS, that are larger than 0.1

NR-SESI - DURATION $(\boldsymbol{\lambda^{\ddagger} = 0.1, \sigma = 0.1})$				**NR-JESI - DURATION** $(\boldsymbol{\lambda^{\ddagger} = 0.1, \sigma = 0.1})$			
Reaction	**MC-LHS**	**NR-DA**	**OHA-DR**	**Reaction**	**MC-LHS**	**NR-DA**	**OHA-DR**
4	28	28	27	4	1	–	1
6	23	24	23	6	0	–	0
11	7	7	7	11	0	–	0
13	16	17	16	13	0	–	0
17	6	5	6	17	1	–	1
21	6	6	6	21	1	–	1

NR-SESI - INTEGRATED RESPONSE $(\boldsymbol{\lambda^{\ddagger} = 0.1, \sigma = 0.1})$				**NR-JESI - INTEGRATED RESPONSE** $(\boldsymbol{\lambda^{\ddagger} = 0.1, \sigma = 0.1})$			
Reaction	**MC-LHS**	**NR-DA**	**OHA-DR**	**Reaction**	**MC-LHS**	**NR-DA**	**OHA-DR**
4	39	39	39	4	1	–	2
6	26	27	27	6	0	–	1
8	5	5	5	8	0	–	1
11	9	9	9	11	0	–	0
13	7	8	8	13	0	–	0

NR-SESI - STRENGTH $(\boldsymbol{\lambda^{\ddagger} = 0.1, \sigma = 0.1})$				**NR-JESI - STRENGTH** $(\boldsymbol{\lambda^{\ddagger} = 0.1, \sigma = 0.1})$			
Reaction	**MC-LHS**	**NR-DA**	**OHA-DR**	**Reaction**	**MC-LHS**	**NR-DA**	**OHA-DR**
4	37	38	36	4	6	–	6
6	15	17	15	6	3	–	6
8	10	10	9	8	3	–	5
11	5	7	4	11	0	–	0
17	6	6	6	17	3	–	4
19	13	14	12	19	2	–	3

Table 5.3: ROSA-based NR-VSI's for the *duration*, *integrated response*, and *strength* of ERK-PP activity in the MAPK signaling cascade, estimated by the three methods considered in this chapter, with $\lambda^{\ddagger} = 0.4$ and $\sigma = 0.7$. Bold reaction numbers indicate NR-SESI or NR-JESI values, obtained by MC-LHS, that are larger than 0.1.

NR-SESI - DURATION $(\lambda^{\ddagger} = 0.4, \sigma = 0.7)$				**NR-JESI - DURATION** $(\lambda^{\ddagger} = 0.4, \sigma = 0.7)$			
Reaction	**MC-LHS**	**NR-DA**	**OHA-DR**	**Reaction**	**MC-LHS**	**NR-DA**	**OHA-DR**
1	7	1	7	1	7	–	6
4	15	25	15	**4**	20	–	22
5	2	6	2	5	6	–	6
6	10	13	10	**6**	14	–	15
8	3	11	3	**8**	12	–	13
11	10	0	11	11	4	–	4
13	7	7	7	13	5	–	5
15	5	11	5	**15**	11	–	11
NR-SESI - INTEGRATED RESPONSE $(\lambda^{\ddagger} = 0.4, \sigma = 0.7)$				**NR-JESI - INTEGRATED RESPONSE** $(\lambda^{\ddagger} = 0.4, \sigma = 0.7)$			
Reaction	**MC-LHS**	**NR-DA**	**OHA-DR**	**Reaction**	**MC-LHS**	**NR-DA**	**OHA-DR**
4	37	28	35	**4**	12	–	13
6	16	17	17	6	9	–	9
8	11	18	11	8	7	–	8
15	5	15	5	15	6	–	6
21	4	4	4	21	5	–	5
NR-SESI - STRENGTH $(\lambda^{\ddagger} = 0.4, \sigma = 0.7)$				**NR-JESI - STRENGTH** $(\lambda^{\ddagger} = 0.4, \sigma = 0.7)$			
Reaction	**MC-LHS**	**NR-DA**	**OHA-DR**	**Reaction**	**MC-LHS**	**NR-DA**	**OHA-DR**
4	34	27	35	**4**	17	–	18
5	3	5	3	5	5	–	5
6	13	16	12	**6**	12	–	12
8	11	15	11	**8**	11	–	12
15	7	15	8	15	9	–	10
21	4	5	4	21	8	–	8

It is clear from the results depicted in Table 5.2 that, for all three response characteristics, the influential reactions exert their influence only singularly, since all NR-JESI values are very small. This indicates that the response functions may be approximately additive within the range of applied fluctuations. It turns out that the NR-SESI's associated with a nearly additive response function can be well estimated by both NR-DA and OHA-DR, as demonstrated by the results in Table 5.2 (and Table 5.6 in the Appendix at the end of this chapter). In this case, NR-DA seems to be the best choice, since it is computationally the most efficient method.

When $\lambda^{\ddagger} = 0.4$ and $\sigma = 0.7$, the number of system evaluations required to estimate the NR-VSI's for the duration, integrated response, and strength, by both ROSA-based OHA-DR and SOSA-based OHA-DR, is about 64%, 55%, and 64% of that required by MC-LHS, respectively. Inspection of the estimated group NR-SESI value $\overline{\sigma}_0$ and group NR-JESI value $\overline{\eta}_0$ reveals that the biochemical factor group $\mathbf{U}_0$ determined by OHA is truly non-influential, since both $\overline{\sigma}_0$ and $\overline{\eta}_0$ are less than 0.1 in all these cases. From the results depicted in Table 5.3 (and Table 5.7 in the Appendix at the end of this chapter), it is clear that NR-DA produces inaccurate results, deeming the use of derivative approximation inappropriate in this case. First of all, NR-DA cannot be used to estimate the NR-JESI values, which can be substantial for large biological or experimental fluctuations. Second, the NR-SESI values estimated by NR-DA differ substantially from those estimated by MC-LHS. For example, the largest difference between the NR-SESI values depicted in Table 5.2 obtained by NR-DA and MC-LHS is 10% for the duration, 10% for the integrated response, and 8% for the strength. On the other hand, OHA-DR consistently provides good results. As a matter of fact, the

Table 5.4: MAPK noise-reduced ROSA results when $\lambda^{\ddagger} = 0.4$ and $\sigma = 0.7$.

No.	Reaction	D	I	S
4	Raf* + Pho1 $\rightleftharpoons$ Raf*-Pho1	•	•	•
6	MEK + Raf* $\rightleftharpoons$ MEK-Raf*	•	•	•
8	MEK-P + Raf* $\rightleftharpoons$ MEK-P-Raf*	•	•	•
11	MEK-PP-Pho2 $\rightarrow$ MEK-P + Pho2	•		
15	ERK-MEK-PP $\rightarrow$ ERK-P + MEK-PP	•		

largest difference between the values depicted in Table 5.2 obtained by OHA-DR and MC-LHS is only 2% for both NR-SESI's and NR-JESI's. Similar conclusions can be made for SOSA by referring to the result summarized in Table 5.7 of the Appendix at the end of this chapter.

As a consequence of the ROSA-based NR-VSI results depicted in Table 5.3, we may conclude that the duration of ERK-PP activity in the MAPK signaling cascade is predominantly influenced by reactions 4, 6, 8, 11, and 15, with reactions 4, 6, and 8 predominantly influencing the integrated response and strength as well. On the other hand, and as a consequence of the SOSA-based NR-VSI results depicted in Table 5.7 of the Appendix at the end of this chapter, we may conclude that the duration of ERK-PP activity in the MAPK signaling cascade is predominantly influenced by molecular species 5, 7, 9, and 16, with species 5, 7, and 9 predominantly influencing the integrated response and strength as well. We summarize these conclusions in Tables 5.4 and 5.5.

Table 5.5: MAPK noise-reduced SOSA results when $\lambda^{\ddagger} = 0.4$ and $\sigma = 0.7$.

No.	Molecular Species	D	I	S
5	Pho1	•	•	•
7	MEK	•	•	•
9	MEK-P	•	•	•
16	ERK-MEK-PP	•		

5.7 Discussion

By comparing the noise-reduced ROSA results summarized in Table 5.4 with the ROSA results depicted in Table 3.2, we conclude that both methods indicate that the strength of ERK-PP activity in the MAPK signaling cascade is mostly influenced by the same reactions, namely reactions 4, 6, and 8. However, in addition to reactions 4 and 6, the noise-reduced ROSA results indicate that reaction 8 influences the timing of ERK-PP activity, whereas, reaction 13 is replaced by reactions 8, 11, and 15 in influencing the duration. It is clear that, by considering experimental variability in the nominal values of rate constants, a higher correlation emerges in the control of the three response characteristics considered in this dissertation, since these characteristics are now commonly influenced by reactions 4, 6, and 8. Reaction 4 is the binding and unbinding of the active version Raf* of the Raf kinase with its inactivator phosphatase Pho1, which is a key step in Raf dephosphorylation. Reactions 6 and 8 both belong to MEK phosphorylation. It has been reported that Raf dephosphorylation and MEK phosphorylation exercise high control on all three characteristics of the signal output in the

MAPK signaling cascade [16], and this is in agreement with our ROSA results based on NR-VSI's. In terms of influential reactions for the duration of ERK-PP activity, reaction 11 is the second step in MEK dephosphorylation, which has been reported to greatly impact the decay time of ERK-PP output [50]. On the other hand, reaction 15 is the single ERK phosphorylation by MEK, which influences the duration only jointly. Inspection of the estimated noise-reduced pairwise-effect sensitivity index[19] reveals that reaction 15 influences the duration of ERK-PP activity mostly with reaction 6, which belongs to MEK phosphorylation. It has been reported that there is a tight connection between ERK phosphorylation and MEK phosphorylation, since ERK is the only known MEK substrate and ERK is phosphorylated when MEK is phosphorylated [112]. This observation is well explained by our noise-reduced ROSA results.

By comparing the noise-reduced SOSA results summarized in Table 5.5 with the ROSA results depicted in Table 3.3, we may also conclude that, by considering experimental variability in the nominal values of the rate constants, a higher correlation emerges in the control of all three response characteristics, since these characteristics are now commonly influenced by molecular species 5, 7, and 9. Species 5 is the phosphatase associated with Raf inactivation. Interestingly, the first drug licensed to act on the MAPK signaling pathway[20] is a Raf kinase inhibitor, which induces anti-proliferative and proapoptotic effects by influencing ERK activity [69]. Molecular species 7 (MEK) and 9 (MEK-P) are the other two species that commonly influence the

[19]This is the noised-reduced version of PESI, defined in Section 3.1.

[20]This is the drug Sorafenib which has been approved for the treatment of primary kidney cancer and advanced primary liver cancer.

three response characteristics. It has been reported that genetically mutated constitutively active MEK is sufficient to cause cellular transformation, which can eventually lead to cancer [72]. In fact, MEK may be easier to target than ERK, since MEK1 is confined to the cytoplasm [112, 113], which makes it a desirable target to block or enhance the response of the MAPK signaling cascade. The results obtained by the noise-reduced SOSA reveal a new influential molecular species for the duration of ERK-PP activity, namely species 16 (ERK-MEK-PP). It has been reported that increased MEK/ERK complex formation contributes to activation of ERK signaling during liver regeneration [114].

In a computational and experimental study, Bhalla *et al.* proposed that differences in the concentrations of MAPK phosphatases, such as protein phosphatase 2A (PP2A) and especially MKP, allow ERK activity to switch between monostable and bistable behavior; concentrations near and below 0.6μM for PP2A or 0.02μM for MKP would allow ERK to exhibit sustained or bistable behavior [115]. However, a simple calculation indicates that the concentration of Pho3 in our model, which acts to dephosphorylate ERK, is approximately 17μM, which would always force ERK to completely adapt back to its original non-phosphorylated form. This brings up issues related to variations in the initial concentrations of different molecular species, which is another commonly studied quantity in sensitivity analysis. In fact, our NR-VSI's can be straightforwardly extended to investigate the influence of different initial component concentrations on response characteristics, under various nominal values of reaction rate constants, or to investigate the influence of different reaction rate constants on the response characteristics, under various nominal values of initial concentrations of molecular species.

To use the noise-reduced sensitivity analysis approach discussed in this chapter, we must specify appropriate values for the standard deviations $\lambda^{\ddagger}$ (or λ) and σ of the underlying biochemical and noise factor perturbations, which are determined by the levels of biological and experimental variations, respectively. A number of quantitative studies have been performed to estimate the magnitudes of biological and experimental variability under different experimental methodologies, based, for example, on two-dimensional gel electrophoresis [39, 42], isobaric tag for relative and absolute quantitation (iTRAQ) [40], and DNA microarrays [41]. In principle, our noise-reduced sensitivity analysis approach can benefit from these studies, which can be used to provide appropriate values for the underlying standard deviations under various experimental conditions. However, it is not always possible to obtain quantitative information about the size of biological and experimental variability. As a consequence, and in most practical situations, we can view the standard deviations $\lambda^{\ddagger}$ (or λ), and σ as user-defined parameters that effectively control the "scale" of sensitivity analysis performed. In particular, the standard deviation σ controls the "range" of probable nominal rate constant values, whereas, the standard deviation $\lambda^{\ddagger}$ (or λ) controls the "size" of parameter fluctuations around a given set of nominal values.

As discussed in Section 5.5 and demonstrated by the simulation results in Section 5.6 and in the Appendix at the end of this chapter, the number of system evaluations required by OHA-based dimensionality reduction decreases linearly as the number of non-influential biochemical factors determined by OHA increases. This number varies for different biochemical reaction systems and different response characteristics. When sensitivity analysis employs the simultaneous use of several response characteristics,

such as the duration, integrated response, and strength of ERK-PP concentration we considered in Section 5.5, the final set of non-influential factors used for dimensionality reduction must be the intersection of all non-influential factors determined by OHA for each individual system response. As a consequence, the final number of non-influential factors may be small and the computational savings achieved by dimensionality reduction may not be significant.

Finally, it is worthwhile noticing that the computational time required by OHA-based factor pre-screening is increased by the time required to solve the polynomial regression required for determining the coefficients in the system response approximation given by Eq. (5.29). This additional time depends on the total number of biochemical and noise factors, which can be substantial for large biochemical reaction systems. On the other hand, the computational time required for solving the system of differential equations given by Eq. (2.2) depends on the size of the biochemical reaction system under consideration and on whether or not the system is stiff. For large biochemical reaction systems with low stiffness, the computational time required for the regression step in OHA may be substantially larger than the time required for solving the underlying differential equations, in which case, the use of MC-LHS may be more preferable than OHA-based dimensionality reduction. In the future, we need to develop new techniques for pre-screening non-influential biochemical factors in order to lower the dimensionality of noise-reduced sensitivity analysis. These techniques should maintain good approximation accuracy of the variance-based sensitivity indices, without requiring a dramatic increase in computational cost for large biochemical reaction systems.

5.8 Appendix

In this appendix, we provide the SOSA-based NR-VSI results for the duration, integrated response, and strength of ERK-PP activity in the MAPK signaling cascade obtained by the three techniques (MC-LHS, NR-DA, and OHA-DR) considered in this chapter. In the first row of Fig. 5.2, we depict the SESI and JESI values for the duration of ERK-PP activity estimated by SOSA-based MC-LHS with standard deviation $\lambda = 0.1$, based on the published and perturbed nominal rate values, respectively. In the second row of Fig. 5.2, we depict the NR-SESI and NR-JESI values for the duration of ERK-PP activity estimated by SOSA-based MC-LHS with standard deviations $\lambda = 0.1$ and $\sigma = 0.7$, based on the published and perturbed nominal rate values, respectively.

In Tables 5.6 and 5.7, we summarize the SOSA-based NR-SESI and NR-JESI values for the duration, integrated response, and strength of ERK-PP, estimated by the three methods considered in this chapter (i.e., MC-LHS, NR-DA, and OHA-DR), with standard deviations $\lambda = 0.1$, $\sigma = 0.1$, in Table 5.6, and $\lambda = 0.4$, $\sigma = 0.7$ in Table 5.7, based on the published nominal values. The results are given in percentages and have been truncated to the nearest integers. To reduce the size of these tables, we depict only the results associated with molecular species whose truncated NR-SESI or NR-JESI values, estimated by MC-LHS, are at least 0.05. Bold species numbers indicate NR-SESI or NR-JESI values, obtained by MC-LHS, that are larger than 0.1, in which case, the corresponding species are deemed to be influential.

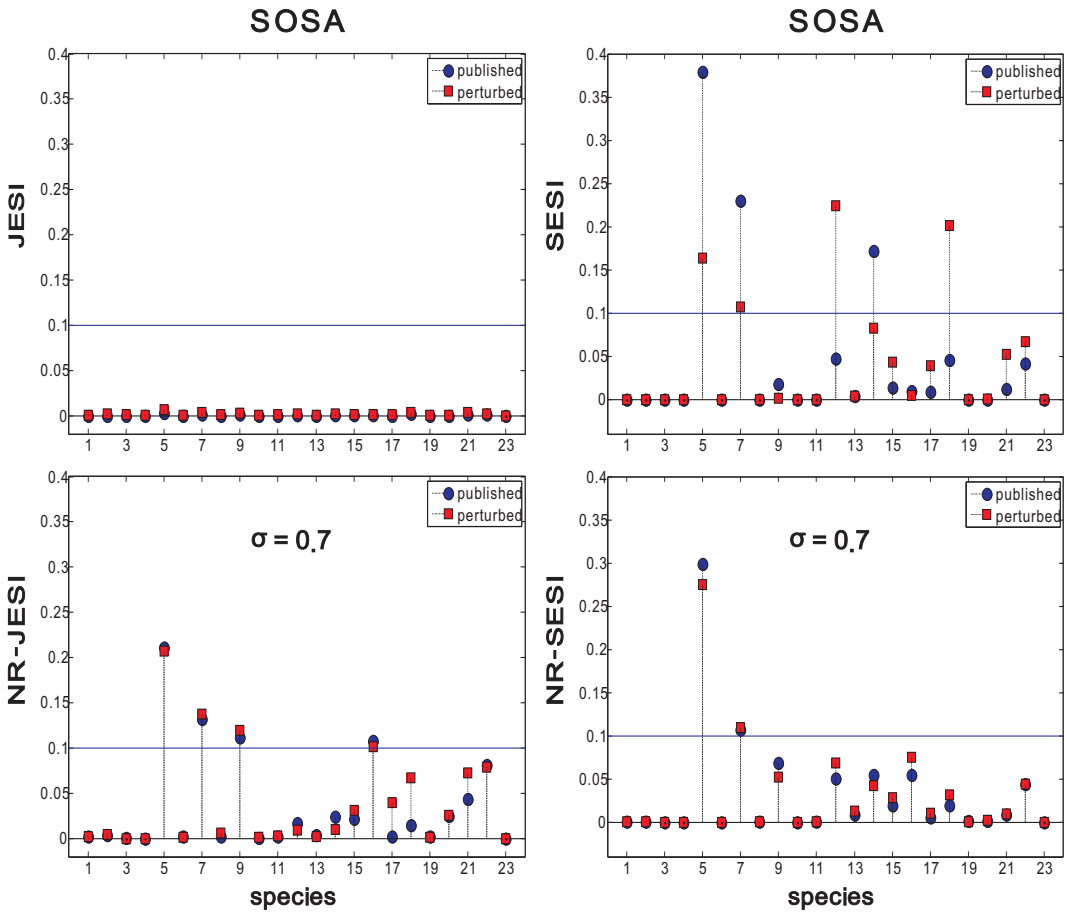

Figure 5.2: SOSA results for the duration of ERK-PP activity in the MAPK signaling cascade based on the published and perturbed nominal rate values with $\lambda = 0.1$.

Table 5.6: SOSA-based NR-VSI's for the *duration*, *integrated response*, and *strength* of ERK-PP activity in MAPK signaling cascade, estimated by the three methods considered in this chapter, with $\lambda = 0.1$ and $\sigma = 0.1$. Bold species numbers indicate NR-SESI or NR-JESI values, obtained by MC-LHS, that are larger than 0.1.

NR-SESI - DURATION $(\lambda^{\ddagger} = 0.1, \sigma = 0.1)$				**NR-JESI - DURATION** $(\lambda^{\ddagger} = 0.1, \sigma = 0.1)$			
Species	**MC-LHS**	**NR-DA**	**OHA-DR**	**Species**	**MC-LHS**	**NR-DA**	**OHA-DR**
5	36	38	37	5	1	–	1
7	21	23	22	7	0	–	0
14	15	17	16	14	0	–	0
18	6	5	6	18	1	–	1
22	5	5	6	22	1	–	1
NR-SESI - INTEGRATED RESPONSE $(\lambda^{\ddagger} = 0.1, \sigma = 0.1)$				**NR-JESI - INTEGRATED RESPONSE** $(\lambda^{\ddagger} = 0.1, \sigma = 0.1)$			
Species	**MC-LHS**	**NR-DA**	**OHA-DR**	**Species**	**MC-LHS**	**NR-DA**	**OHA-DR**
5	47	47	48	5	1	–	2
7	22	23	22	7	1	–	1
9	9	9	9	9	1	–	1
14	10	11	10	14	0	–	0
NR-SESI - STRENGTH $(\lambda^{\ddagger} = 0.1, \sigma = 0.1)$				**NR-JESI - STRENGTH** $(\lambda^{\ddagger} = 0.1, \sigma = 0.1)$			
Species	**MC-LHS**	**NR-DA**	**OHA-DR**	**Species**	**MC-LHS**	**NR-DA**	**OHA-DR**
5	40	41	38	5	9	–	9
7	12	13	10	7	4	–	5
9	20	25	22	9	6	–	5
21	8	7	7	21	2	–	2

Table 5.7: SOSA-based NR-VSI's for the *duration*, *integrated response*, and *strength* of ERK-PP activity in MAPK signaling cascade, estimated by the three methods considered in this chapter, with $\lambda = 0.4$ and $\sigma = 0.7$. Bold species numbers indicate NR-SESI or NR-JESI values, obtained by MC-LHS, that are larger than 0.1.

NR-SESI - DURATION $(\boldsymbol{\lambda^{\ddagger} = 0.4, \sigma = 0.7})$				**NR-JESI - DURATION** $(\boldsymbol{\lambda^{\ddagger} = 0.4, \sigma = 0.7})$			
Species	**MC-LHS**	**NR-DA**	**OHA-DR**	**Species**	**MC-LHS**	**NR-DA**	**OHA-DR**
5	24	31	24	**5**	25	–	25
7	10	7	10	**7**	16	–	16
9	5	28	4	**9**	14	–	14
12	8	2	9	12	4	–	4
14	8	14	8	14	4	–	4
16	5	2	5	**16**	11	–	11
22	2	3	2	22	8	–	8
NR-SESI - INTEGRATED RESPONSE $(\boldsymbol{\lambda^{\ddagger} = 0.4, \sigma = 0.7})$				**NR-JESI - INTEGRATED RESPONSE** $(\boldsymbol{\lambda^{\ddagger} = 0.4, \sigma = 0.7})$			
Species	**MC-LHS**	**NR-DA**	**OHA-DR**	**Species**	**MC-LHS**	**NR-DA**	**OHA-DR**
5	47	38	47	**5**	13	–	13
7	13	28	15	7	8	–	9
9	10	27	11	9	7	–	8
16	4	1	4	16	5	–	6
22	3	2	3	22	5	–	5
NR-SESI - STRENGTH $(\boldsymbol{\lambda^{\ddagger} = 0.4, \sigma = 0.7})$				**NR-JESI - STRENGTH** $(\boldsymbol{\lambda^{\ddagger} = 0.4, \sigma = 0.7})$			
Species	**MC-LHS**	**NR-DA**	**OHA-DR**	**Species**	**MC-LHS**	**NR-DA**	**OHA-DR**
5	43	38	44	**5**	18	–	19
7	11	27	12	**7**	11	–	11
9	10	29	10	**9**	10	–	11
16	6	1	6	16	9	–	9
22	3	1	3	22	7	–	7

Chapter 6

Conclusions

In this dissertation, we presented a thermodynamically consistent variance-based approach for the sensitivity analysis of biochemical reaction systems. We developed a probabilistic model for perturbing the reaction rate constants based on the classical Eyring-Polanyi equations of chemical kinetics. This model guarantees that the perturbed rate constants automatically satisfy the Wegscheider conditions imposed by thermodynamics. As a consequence, and in sharp contrast to previously proposed techniques, the sensitivity analysis approach we introduced in this dissertation leads to thermodynamically consistent results. Compared with the most commonly used derivative-based sensitivity analysis approach, variance-based sensitivity analysis can easily accommodate appreciable parameter variations and allows for a systematic investigation of interactions among different system components. Besides, variance-based sensitivity analysis is not limited to additive system responses, and is able to treat groups of factors as if they were single ones. Our numerical results demonstrate that the proposed method is very appealing, since it can produce rich information about

the sensitivity properties of a biochemical reaction system and can lead to rigorous and systematic interpretation of the results. However, variance-based sensitivity analysis is computationally intensive due to the need for a large number of system evaluations in a Monte Carlo framework. This problem becomes very serious for large biochemical reaction systems, especially when the differential equations governing the system are stiff.

Motivated by the previous computational issue, we discussed four techniques that one can use to analytically approximate the second-order sensitivity indices associated with the proposed variance-based sensitivity analysis methodology. We highlighted important theoretical, numerical, and computational aspects of each method, in an attempt to provide a comprehensive understanding of the advantages and disadvantages of each technique. Our simulation results, based on a mathematical model of the MAPK signaling cascade, clearly demonstrate the inferiority of second-order derivative-based sensitivity analysis at moderate to high levels of uncertainty. It also shows the superiority of OHA, which is constructed by truncating the ANOVA-HDMR of the response function of a biochemical reaction system and by approximating the first- and second-order ANOVA-HDMR component functions using orthonormal Hermite polynomials. All four approximation techniques are orders of magnitude faster than Monte Carlo estimation, and make the variance-based sensitivity analysis approach more practical for large biochemical reaction systems.

As it is true in the case of derivative-based sensitivity analysis techniques, the application of the variance-based sensitivity analysis approach discussed in this thesis requires specifying the nominal values of the kinetic parameters. As a consequence,

the sensitivity analysis results may depend on the particular choice of the nominal parameter values. To address this issue, we proposed an extended version of our thermodynamically consistent variance-based approach, which effectively reduces the experimental uncertainty associated with the choice of nominal kinetic parameter values. Our simulation results clearly demonstrated the superiority of the noise-reduced variance-based sensitivity indices for quantifying the average relative importance of different biochemical factors over a broad range of nominal rate values, thus leading to biological conclusions that are less sensitive to the uncertainties in choosing these values. To reduce the computational cost of Monte Carlo estimation when evaluating noise-reduced variance-based sensitivity indices, we discussed two numerical methods based on derivative approximation and OHA-based dimensionality reduction, respectively. Although derivative approximation is a very fast method, its applicability is limited to nearly additive response functions. On the other hand, OHA-based dimensionality reduction is applicable to more general response functions. However, the computational cost of this method depends on a polynomial regression step, which can be inefficient when dealing with large biochemical reaction systems, and on the number of non-influential biochemical factors detected by pre-screening. To further improve the noise-reduced variance-based sensitivity analysis approach, future research must be focused on developing numerically accurate and computationally efficient techniques for pre-screening non-influential factors in large biochemical reaction systems.

The probabilistic sensitivity analysis techniques discussed in this thesis were demonstrated using a biologically relevant example, namely the MAPK signaling cascade, in terms of three response characteristics with established biological significance. The sen-

sitivity analysis results identified only a few reactions and molecular species as being influential on the duration, integrated response, and strength of the output ERK-PP activity. During this doctoral research, we have also employed a number of other well-known cellular signaling pathways, such as the epidermal growth factor receptor (EGFR) signaling pathway [116], the EGF-ERK signaling pathway [45], and the Janus kinase-signal transducers and activators of transcription (JAK-STAT) signaling pathway [117], to test the sensitivity analysis approaches we considered in this thesis. The results we obtained clearly indicate that the biological behavior of cell signaling depends only on a handful of biochemical factors. This property makes sensitivity analysis a desirable tool for simplifying biochemical reaction system models, for identifying influential targets in system-based drug design, and for estimating system model parameters more efficiently and accurately.

An important assumption made in this dissertation is that input biochemical factors are independent of each other. As we discussed in Section 2.4, it is not easy to justify the mutual independence between the standard chemical potentials of activated complexes associated with closely related reactions, and the mutual independence between the standard chemical potentials of correlated molecular species, especially those with common components. More accurate modeling to reflect relationships between these biochemical factors needs biophysical knowledge on how biochemical factors depend on each other. Another reason for the independence assumption is that correlated input samples are more laborious to generate. Moreover, the sample size needed to compute sensitivity measures for correlated samples is much higher than that for independent ones [23, 27]. A useful trick to circumvent the use of correlated samples is to treat

dependencies as being produced by a known transformation applied on statistically independent noise terms [23], a process known as whitening. We believe that future work should be done to develop appropriate probabilistic models and sensitivity measures that can efficiently handle correlated biochemical factors.

The sensitivity analysis approaches discussed in this dissertation were applied to deterministic chemical kinetics based on the mass action rate law. Using this type of kinetics requires the assumption that the number of molecular species is large and represents a macroscopic view of a biochemical reaction system. However, the number of molecular species involved in some gene networks and cellular signaling pathways can be quite small, and stochastic effects may dominate the system behavior [118,119]. In this case, biochemical reaction systems are usually modeled by using the chemical master equation [120] or, under certain conditions, by employing the chemical Langevin equation [121]. We can extend the variance-based sensitivity analysis approaches discussed in this thesis to this case by evaluating system responses using Gillespie's stochastic simulation algorithm (SSA) [122, 123]. However, the computational cost involved would be prohibitively expensive even for small biochemical reaction systems. The approximation techniques discussed in this dissertation can be applied to reduce the computational burden.

Bibliography

[1] H. Kitano, "Systems biology: A brief overview," *Science*, vol. 295, pp. 1662–1664, 2002.

[2] ——, "Computational systems biology," *Nature*, vol. 420, pp. 206–210, 2002.

[3] I. Nestorov, "Whole body pharmacokinetic models," *Clinical Pharmacokinetics*, vol. 42, no. 10, pp. 883–908, 2003.

[4] A. Varma, M. Morbidelli, and H. Wu, *Parametric Sensitivity in Chemical Systems*. Cambridge, United Kingdom: Cambridge University Press, 1999.

[5] A. Saltelli, M. Ratto, S. Tarantola, and F. Campolongo, "Sensitivity analysis for chemical models," *Chemical Reviews*, vol. 105, pp. 2811–2827, 2005.

[6] H. Xu and S. Rahman, "Decomposition methods for structural reliability analysis," *Probabilistic Engineering Mechanics*, vol. 20, pp. 239–250, 2005.

[7] H. Liu, W. Chen, and A. Sudjianto, "Relative entropy based method for probabilistic sensitivity analysis in engineering design," *Journal of Mechanical Design*, vol. 128, pp. 326–336, 2006.

[8] A. Saltelli and S. Tarantola, "On the relative importance of input factors in mathematical models: Saftey assessment for nuclear waste disposal," *Journal of the American Statistical Association*, vol. 97, no. 459, pp. 702–709, 2002.

[9] A. Fassò, E. Esposito, E. Porcu, A. P. Reverberi, and F. Vegliò, "Statistical sensitivity analysis of packed column reactors for contaminated wastewater," *Environmetrics*, vol. 14, no. 8, pp. 743–759, 2003.

[10] D. Krewski, Y. Wang, S. Bartlett, and K. Krishnan, "Uncertainty, variability, and sensitivity in physiological pharmacokinetic models," *Journal of Biopharmaceutical Statistics*, vol. 5, no. 3, pp. 245–271, 1995.

[11] H. Rabitz, M. Kramer, and D. Dacol, "Sensitivity analysis in chemical kinetics," *Annual Review of Physical Chemistry*, vol. 34, pp. 419–461, 1983.

[12] J. Zádor, I. G. Zsély, and T. Turányi, "Local and global uncertainty analysis of complex chemical kinetic systems," *Reliability Engineering and System Safety*, vol. 91, pp. 1232–1240, 2006.

[13] E. E. Leamer, "Sensitivity analysis would help," in *Modelling Economic Series: Readings in Econometric Methodology*, C. W. J. Granger, Ed. Oxford University Press, 1991.

[14] F. Campolongo, J. Cariboni, and W. Schoutens, "The importance of jumps in pricing European options," *Reliability Engineering and System Safety*, vol. 91, pp. 1148–1154, 2006.

[15] D. Baur, J. Cariboni, and F. Campolongo, "Global sensitivity analysis for latent factor credit risk models," *International Journal of Risk Assessment and Management*, vol. 11, pp. 281–298, 2009.

[16] J. J. Hornberg, B. Binder, F. J. Bruggeman, B. Schoeberl, R. Heinrich, and H. V. Westerhoff, "Control of MAPK signalling: From complexity to what really matters," *Oncogene*, vol. 24, pp. 5533–5542, 2005.

[17] G. Liu, M. T. Swihart, and S. Neelamegham, "Sensitivity, principal component and flux analysis applied to signal transduction: The case of epidermal growth factor mediated signaling," *Bioinformatics*, vol. 21, no. 7, pp. 1194–1202, 2005.

[18] D. Hu and J.-M. Yuan, "Time-dependent sensitivity analysis of biological networks: Coupled MAPK and PI3K signal transduction pathways," *Journal of Physical Chemistry A*, vol. 110, pp. 5361–5370, 2006.

[19] A. Mahdavi, R. E. Davey, P. Bhola, T. Yin, and P. W. Zandstra, "Sensitivity analysis of intracellular signaling pathway kinetics predicts targets for stem cell fate control," *PLoS Computational Biology*, vol. 3, no. 7, pp. 1257–1267, 2007.

[20] X.-J. Feng, S. Hooshangi, D. Chen, G. Li, R. Weiss, and H. Rabitz, "Optimizing genetic circuits by global sensitivity analysis," *Biophysical Journal*, vol. 87, pp. 2195–2202, 2004.

[21] J.-C. Leloup and A. Goldbeter, "Modeling the mammalian circadian clock: Sensitivity analysis and multiplicity of oscillatory mechanisms," *Journal of Theoretical Biology*, vol. 230, pp. 541–562, 2004.

[22] H. V. Westerhoff, "Systems biology: New paradigms for cell biology and drug design," in *Systems Biology: Applications and Perspectives*, P. Bringmann, E. C. Butcher, G. Parry, and B. Weiss, Eds. Springer, 2007, pp. 45–67.

[23] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola, *Global Sensitivity Analysis: The Primer*. Chichester, England: John Wiley, 2008.

[24] S. M. Stephens and J. Rung, "Advances in systems biology: measurement, modeling and representation," *Current Opinion in Drug Discovery and Development*, vol. 9, no. 2, pp. 240–250, 2006.

[25] I. M. Sobol', "Sensitivity estimates for nonlinear mathematical models," *Mathematical Modeling and Computational Experiment*, vol. 1, no. 4, pp. 407–414, 1993.

[26] ——, "Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates," *Mathematics and Computers in Simulation*, vol. 55, pp. 271–280, 2001.

[27] A. Saltelli, S. Tarantola, F. Campolongo, and M. Ratto, *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. Chichester, England: John Wiley, 2004.

[28] N. Lüdtke, S. Panzeri, M. Brown, D. S. Broomhead, J. Knowles, M. A. Montemurro, and D. B. Kell, "Information-theoretic sensitivity analysis: a general

method for credit assignment in complex networks," *Journal of the Royal Society Interface*, vol. 5, no. 19, pp. 223–235, 2008.

[29] T. Schürmann, "Bias analysis in entropy estimation," *Journal of Physics A: Mathematical and General*, vol. 37, no. 27, pp. L295–L301, 2004.

[30] M. Ederer and E. D. Gilles, "Thermodynamically feasible kinetic models of reaction networks," *Biophysical Journal*, vol. 92, pp. 1846–1857, 2007.

[31] A. Saltelli, "Making best use of model evaluations to compute sensitivity indices," *Computer Physics Communications*, vol. 145, pp. 280–297, 2002.

[32] H.-X. Zhang, W. P. Dempsey, and J. Goutsias, "Probabilistic sensitivity analysis of biochemical reaction systems," *Journal of Chemical Physics*, vol. 131, pp. 1–20, 2009.

[33] J. S. Liu, *Monte Carlo Strategies in Scientific Computing*. New York: Springer, 2001.

[34] J. C. Helton and F. J. Davis, "Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems," *Reliability Engineering and System Safety*, vol. 81, pp. 23–69, 2003.

[35] H. Rabitz, Ö. F. Alis, J. Shorter, and K. Shim, "Efficient input-output model representations," *Computer Physics Communications*, vol. 117, pp. 11–20, 1999.

[36] H. Rabitz and Ö. F. Alis, "General foundations of high-dimensional model representations," *Journal of Mathematical Chemistry*, vol. 25, pp. 197–233, 1999.

[37] G. Li, C. Rosenthal, and H. Rabitz, "High dimensional model representations," *Journal of Physical Chemistry A*, vol. 105, pp. 7765–7777, 2001.

[38] K. Aggarwal and K. H. Lee, "Functional genomics and proteomics as a foundation for systems biology," *Briefings in Functional Genomics and Proteomics*, vol. 2, no. 3, pp. 175–184, 2003.

[39] M. P. Molloy, E. E. Brzezinski, J. Hang, M. T. McDowell, and R. A. VanBogelen, "Overcoming technical variation and biological variation in quantitative proteomics," *Proteomics*, vol. 3, pp. 1912–1919, 2003.

[40] C. S. Gan, P. K. Chong, T. K. Pham, and P. C. Wright, "Technical, experimental, and biological variations in isobaric tags for relative and absolute quantitation (iTRAQ)," *Journal of Proteome Research*, vol. 6, pp. 821–827, 2007.

[41] S. O. Zakharkin, K. Kim, T. Mehta, L. Chen, S. Barnes, K. E. Scheirer, R. S. Parrish, D. B. Allison, and G. P. Page, "Sources of variation in Affymetrix microarray experiments," *BMC Bioinformatics*, vol. 6, p. 241, 2005.

[42] A. M. Bland, M. G. Janech, J. S. Almeida, and J. M. Arthur, "Sources of variability among replicate samples separated by two-dimensional gel electrophoresis," *Journal of Biomolecular Technics*, vol. 21, no. 1, pp. 3–8, 2010.

[43] R. Heinrich and S. Schuster, *The Regulation of Cellular Systems*. New York City, New York: Chapman & Hall, 1996.

[44] L. Chang and M. Karin, "Mammalian MAP kinase signalling cascades," *Nature*, vol. 410, pp. 37–40, 2001.

[45] B. Schoeberl, C. Eichler-Jonsson, E. D. Gilles, and G. Müller, "Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors," *Nature Biotechnology*, vol. 20, pp. 370–375, 2002.

[46] M. Kleijn and C. G. Proud, "The regulation of protein synthesis and translation factors by CD3 and CD28 in human primary T lymphocytes," *BMC Biochemistry*, vol. 3, p. 11, 2002.

[47] L. O. Murphy, J. P. MacKeigan, and J. Blenis, "A network of immediate early gene products propagates subtle differences in mitogen-activated protein kinase signal amplitude and duration," *Molecular and Cellular Biology*, vol. 24, no. 1, pp. 144–153, 2004.

[48] C. J. Marshall, "Specificity of receptor tyrosine kinase signaling: Transient versus sustained extracellular signal-regulated kinase activation," *Cell*, vol. 80, pp. 179–185, 1995.

[49] L. O. Murphy, S. Smith, R.-H. Chen, D. C. Fingar, and J. Blenis, "Molecular interpretation of ERK signal duration by immediate early gene products," *Nature Cell Biology*, vol. 4, pp. 556–564, 2002.

[50] K. Mayawala, C. A. Gelmi, and J. S. Edwards, "MAPK cascade possesses decoupled controllability of signal amplification and duration," *Biophysical Journal*, pp. L01–L02, 2004.

[51] M. Ebisuya, K. Kondoh, and E. Nishida, "The duration, magnitude and compart-

mentalization of ERK MAP kinase activity: Mechanisms for providing signaling specificity," *Journal of Cell Science*, vol. 118, pp. 2997–3002, 2005.

[52] R. M. Tombes, K. L. Auer, R. Mikkelsen, K. Valerie, M. P. Wymann, C. J. Marshall, M. McMahon, and P. Dent, "The mitogen-activated protein (MAP) kinase cascade can either stimulate or inhibit DNA synthesis in primary cultures of rat hepatocytes depending upon whether its activation is acute/phasic or chronic," *The Biochemical Journal*, vol. 330, pp. 1451–1460, 1998.

[53] A. R. Asthagiri, C. A. Reinhart, A. F. Horwitz, and D. A. Lauffenburger, "The role of transient ERK2 signals in fibronectin- and insulin-mediated DNA synthesis," *Journal of Cell Science*, vol. 113, pp. 4499–4510, 2000.

[54] S. Traverse, K. Seedorf, H. Paterson, C. J. Marshall, P. Cohen, and A. Ullrich, "EGF triggers neural differentiation of PC12 cells that overexpress the EGF receptor," *Current Biology*, vol. 4, no. 8, pp. 694–701, 1994.

[55] R. S. Berry, S. A. Rice, and J. Ross, *Physical Chemistry*, 2nd ed. New York City, New York: Oxford University Press, 2000.

[56] X.-S. Xie and H.-P. Lu, "Single-molecule enzymology," *International Journal for Numerical Methods in Engineering*, vol. 274, pp. 15 967–15 970, 1999.

[57] X.-S. Xie, "Single-molecule approach to dispersed kinetics and dynamic disorder: Probing conformational fluctuation and enzymatic dynamics," *Journal of Chemical Physics*, vol. 117, pp. 1–9, 2002.

[58] J. M. Berg, J. L. Tymoczko, and L. Stryer, *Biochemistry*, 6th ed. New York City, New York: W. H. Freeman, 2006.

[59] M. O. Vlad, G. Cerofolini, P. Oefner, and J. Ross, "Random activation energy model and disordered kinetics, from static to dynamic disorder," *Journal of Physical Chemistry B*, vol. 109, pp. 21 241–21 257, 2005.

[60] T. Grönholm and A. Annila, "Natural distribution," *Mathematical Biosciences*, vol. 210, pp. 659–667, 2007.

[61] W. Liebermeister and E. Klipp, "Biochemical networks with uncertain parameters," *IEE Proceedings Systems Biology*, vol. 152, no. 3, pp. 97–107, 2005.

[62] J. Schaber, W. Liebermeister, and E. Klipp, "Nested uncertainties in biochemical models," *IET Systems Biology*, vol. 3, no. 1, pp. 1–9, 2009.

[63] T. Homma and A. Saltelli, "Importance measures in global sensitivity amalysis of nonlinear models," *Reliability Engineering and System Safety*, vol. 52, pp. 1–17, 1996.

[64] A. Saltelli, S. Tarantola, and F. Campolongo, "Sensitivity analysis as an ingredient of modeling," *Statistical Science*, vol. 15, no. 4, pp. 377–395, 2000.

[65] R. L. Iman, "A matrix-based approach to uncertainty and sensitivity analysis for fault trees," *Risk Analysis*, vol. 7, no. 1, pp. 21–33, 1987.

[66] M. D. McKay, W. J. Conover, and R. J. Beckman, "A comparison of three methods for selectioin values of input variables in the analysis of output from a computer code," *Technometrics*, vol. 21, no. 2, pp. 239–245, 1979.

[67] M. Stein, "Large sample properties of simulations using Latin hypercube sampling," *Technometrics*, vol. 29, no. 2, pp. 143–151, 1987.

[68] J. C. Helton, J. D. Johnson, C. J. Sallaberry, and C. B. Storlie, "Survey of sampling-based methods for uncertainty and sensitivity analysis," *Reliability Engineering and System Safety*, vol. 91, pp. 1175–1209, 2006.

[69] S. M. Wilhelm, L. Adnane, P. Newell, A. Villanueva, J. M. Llovet, and M. Lynch, "Preclinical overview of sorafenib, a multikinase inhibitor that targets both Raf and VEGF and PDGF receptor tyrosine kinase signaling," *Molecular Cancer Therapeutics*, vol. 7, no. 10, pp. 3129–3140, 2008.

[70] G. Bollag, P. Hirth, J. Tsai, J. Zhang, P. N. Ibrahim, H. Cho, W. Spevak, C. Zhang, Y. Zhang, G. Habets, E. A. Burton, B. Wong, G. Tsang, B. L. West, B. Powell, R. Shellooe, A. Marimuthu, H. Nguyen, K. Y. J. Zhang, D. R. Artis, J. Schlessinger, F. Su, B. Higgins, R. Iyer, and K. D'Andrea, "Clinical efficacy of a RAF inhibitor needs broad target blockade in braf-mutant melanoma," *Nature*, vol. 467, pp. 596–599, 2010.

[71] E. Sala, L. Mologni, S. Truffa, C. Gaetano, G. E. Bollag, and C. Gambacorti-Passerini, "BRAF silencing by short hairpin RNA or chemical blockade by PLX4032 leads to different responses in melanoma and thyroid carcinoma cells," *Molecular Cancer Research*, vol. 6, no. 5, pp. 751–759, 2008.

[72] S. J. Mansour, W. T. Matten, A. S. Hermann, J. M. Candia, S. Rong, K. Fukasawa, G. F. Vande Woude, and N. G. Ahn, "Transformation of mammalian cells

by constitutively active MAP kinase kinase," *Science*, vol. 265, pp. 966–970, 1994.

[73] K. Orth, L. E. Palmer, Z. Q. Bao, S. Stewart, A. E. Rudolph, J. B. Bliska, and J. E. Dixon, "Inhibition of the mitogen-activated protein kinase kinase superfamily by a *yersinia* effector," *Science*, vol. 285, no. 5435, pp. 1920–1923, 1999.

[74] I. M. Sobol', "Theorems and examples on high-dimensional model representation," *Reliability Engineering and System Safety*, vol. 79, pp. 187–193, 2003.

[75] J. E. Oakley and A. O'Hagan, "Probabilistic sensitivity analysis of complex models: a Bayesian approach," *Journal of the Royal Statistical Society B*, vol. 66, no. 3, pp. 751–769, 2004.

[76] M. Ratto, A. Pagano, and P. Young, "State dependent parameter metamodelling and sensitivity analysis," *Computer Physics Communications*, vol. 177, pp. 863–876, 2007.

[77] C. B. Storlie and J. C. Helton, "Multiple predictor smoothing methods for sensitivity analysis: Description of techniques," *Reliability Engineering and System Safety*, vol. 93, pp. 28–54, 2008.

[78] C. B. Storlie, L. P. Swiler, J. C. Helton, and C. J. Sallaberry, "Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models," *Reliability Engineering and System Safety*, vol. 94, pp. 1735–1763, 2009.

[79] W. Chen, R. Jin, and A. Sudjianto, "Analytical variance-based global sensitivity analysis in simulation-based design under uncertainty," *ASME Journal of Mechanical Design*, vol. 127, no. 5, pp. 875–886, 2005.

[80] G. E. B. Archer, A. Saltelli, and I. M. Sobol, "Sensitivity measures, anova-like techniques and the use of bootstrap," *Journal of Statistcal Computation and Simulation*, vol. 58, pp. 99–120, 1997.

[81] A. B. Owen, "Latin supercube sampling for very high-dimensional simulations," *ACM Transactions on Modeling and Computer Simulation*, vol. 8, pp. 71–102, 1998.

[82] D. G. Cacuci, *Sensitivity and Uncertainty Analysis*. Boca Raton: Chapman & Hall/CRC, 2003, vol. I. Theory.

[83] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis, Third Edition*. New York City, New York: John Wiley & Sons, Inc., 2001.

[84] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, 3rd ed. New York: Cambridge University Press, 2007.

[85] H. Xu and S. Rahman, "A generalized dimension-reduction method for multidimensional integration in stochastic mechanics," *International Journal for Numerical Methods in Engineering*, vol. 61, pp. 1992–2019, 2004.

[86] Ö. F. Aliş and H. Rabitz, "Efficient implementation of high dimensional model representations," *Journal of Mathematical Chemistry*, vol. 29, pp. 127–142, 2001.

[87] G. Li, S.-W. Wang, and H. Rabitz, "Practical approaches to construct RS-HDMR component functions," *Journal of Physical Chemistry A*, vol. 106, pp. 8721–8733, 2002.

[88] S.-W. Wang, P. G. Georgopoulos, G. Li, and H. Rabitz, "Random sampling-high dimensional model representation (RS-HDMR) with nonuniform distributed variables: Application to an integrated multimedia/multipathway exposure and dose model for trichloroethylene," *Journal of Physical Chemistry A*, vol. 107, pp. 4707–4716, 2003.

[89] S.-K. Choi, R. V. Grandhi, R. A. Canfield, and C. L. Pettit, "Polynomial chaos expansion with Latin hypercube sampling for estimating response variability," *AIAA Journal*, vol. 42, pp. 1191–1198, 2004.

[90] B. Sudret, "Global sensitivity analysis using polynomial chaos expansions," *Reliability Engineering and System Safety*, vol. 93, pp. 964–979, 2008.

[91] T. Crestaux, O. L. Maître, and J.-M. Martinez, "Polynomial chaos expansion for sensitivity analysis," *Reliability Engineering and System Safety*, vol. 94, pp. 1161–1172, 2009.

[92] E. Castillo, N. Sánchez-Maroño, A. Alonso-Betanzos, and C. Castillo, "Functional network topology learning and sensitivity analysis based on ANOVA decomposition," *Neural Computation*, vol. 19, pp. 231–257, 2007.

[93] H.-X. Zhang and J. Goutsias, "A comparison of approximation techniques for variance-based sensitivity analysis of biochemical reaction systems," *BMC Bioinformatics*, vol. 11, p. 246, 2010.

[94] G. Li, J. Hu, S.-W. Wang, P. G. Georgopoulos, J. Schoendorf, and H. Rabitz, "Random sampling-high dimensional model represenattion (RS-HDMR) and orthogonality of its different order component functions," *Journal of Physical Chemistry A*, vol. 110, pp. 2474–2485, 2006.

[95] Z. Zi, K. H. Cho, M. H. Sung, X. Xia, J. Zheng, and Z. Sun, "In silico identification of the key components and steps in IFN-gamma induced JAK-STAT signaling pathway," *FEBS Letters*, vol. 579, pp. 1101–1108, 2005.

[96] K. S. Brown and J. P. Sethna, "Statistical mechanical approaches to models with many poorly known parameters," *Physical Review E*, vol. 68, no. 021904, 2003.

[97] R. N. Gutenkunst, J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers, and J. P. Sethna, "Universally sloppy parameter sensitivities in systems biology models," *PLoS Computational Biology*, vol. 3, no. 10, e189, 2007.

[98] S. Bandara, J. P. Schlöder, R. Eils, H. G. Bock, , and T. Meyer, "Optimal experimental design for parameter estimation of a cell signaling model," *PLoS Computational Biology*, vol. 5, no. 11, e1000558, 2009.

[99] A. P. Minton, "How can biochemical reactions within cells differ from those in test tubes?" *Journal of Cell Science*, vol. 119, pp. 2863–2869, 2006.

[100] H. Motulsky, *Intuitive Biostatistics: A Nonmathematical Guide to Statistical Thinking*. New York City, New York: Oxford University Press, 1995.

[101] A. Kiparissides, S. S. Kucherenko, A. Mantalaris, and E. N. Pistikopoulos, "Global sensitivity analysis challenges in biological systems modeling," *Industrial and Engineering Chemical Research*, vol. 48, pp. 7168–7180, 2009.

[102] A. Saltelli, K. Chan, and E. M. Scott, *Sensitivity Analysis*. New York: Wiley, 2000.

[103] M. D. Morris, "Factorial sampling plans for preliminary computational experiments," *Technometrics*, vol. 33, pp. 161–174, 1991.

[104] F. Campolongo, J. Cariboni, and A. Saltelli, "An effective screening design for sensitivity analysis of large models," *Environmental Modelling and Software*, vol. 22, pp. 1509–1518, 2007.

[105] I. M. Sobol and S. Kucherenko, "Derivative based global sensitivity measures and their link with global sensitivity indices," *Mathematics and Computers in Simulation*, vol. 79, pp. 3009–3017, 2009.

[106] S. Kucherenko, M. Rodriguez-Fernandez, C. Pantelides, and N. Shah, "Monte Carlo evaluation of derivative-based global sensitivity measures," *Reliability Engineering and System Safety*, vol. 94, pp. 1135–1148, 2009.

[107] S. M. Dunn, A. Constantinides, and P. V. Moghe, *Numerical Methods in Biomedical Engineering*. San Diego, California: Elsevier Academic Press, 2006.

[108] U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits*. London, United Kingdom: Chapman & Hall, 2006.

[109] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*. Boca Raton, Florida: Chapman & Hall/CRC, 2004.

[110] W. G. Jenkinson and J. Goutsias, "On constructing thermodynamically valid parameterizations of biochemical systems," *Proceedings of the IEEE International Confonference on Bioinformatics and Bioengineering (BIBE)*, pp. 214–219, 2010.

[111] H.-X. Zhang and J. Goutsias, "Reducing experimental variability in variance-based sensitivity analysis of biochemical reaction systems," *Journal of Chemical Physics*, submitted.

[112] G. Pearson, F. Robinson, T. B. Gibson, B. Xu, M. Karandikar, K. Berman, and M. H. Cobb, "Mitogen-activated protein (MAP) kinase pathways: Regulation and physiological functions," *Endocrine Reviews*, vol. 22, no. 2, pp. 153–183, 2001.

[113] W. Kolch, "Meaningful relationships: The regulation of the Ras/Raf/MEK/ERK pathway by protein interactions," *Biochemical Journal*, vol. 351, pp. 289–305, 2000.

[114] M. Won, K. A. Park, H. S. Byun, Y.-R. Kim, B. L. Choi, J. H. Hong, J. Park, J. H. Seok, Y.-H. Lee, C.-H. Cho, I. S. Song, Y. K. Kim, H.-M. Shen, and G. M. Hur, "Protein kinase SGK1 enhances MEK/ERK complex formation through the phosphorylation of ERK2: Implication for the positive regulatory role of SGK1

on the ERK function during liver regeneration," *Journal of Hepatology*, vol. 51, no. 1, pp. 67–76, 2009.

[115] U. S. Bhalla, P. T. Ram, and R. Iyengar, "MAP kinase phosphatase as a locus of flexibility in a mitogen-activated protein kinase signaling network," *Science*, vol. 297, pp. 1018–1023, 2002.

[116] B. N. Kholodenko, O. V. Demin, G. Moehren, and J. B. Hoek, "Quantification of short term signaling by the epidermal growth factor receptor," *Journal of Biological Chemistry*, vol. 274, no. 42, pp. 30 169–30 181, 1999.

[117] S. Yamada, S. Shiono, A. Joo, and A. Yoshimura, "Control mechanism of JAK/STAT signal transduction pathway," *FEBS Letters*, vol. 534, pp. 190–196, 2003.

[118] J. Paulsson, "Summing up the noise in gene networks," *Nature*, vol. 427, pp. 415–418, 2004.

[119] A. Colman-Lerner, A. Gordon, E. Serra, T. Chin, O. Resnekov, D. Endy, C. G. Pesce, and R. Brent, "Regulated cell-to-cell variation in a cell-fate decision system," *Nature*, vol. 437, pp. 699–706, 2005.

[120] D. T. Gillespie, "A rigorous derivation of the chemical master equation," *Physica A*, vol. 188, pp. 404–425, 1992.

[121] ——, "The chemical Langevin equation," *Journal of Chemical Physics*, vol. 113, pp. 297–306, 2000.

[122] ——, "Exact stochastic simulation of coupled chemical reactions," *Journal of Physical Chemistry*, vol. 81, no. 25, pp. 2340–2361, 1977.

[123] A. Degasperi and S. Gilmore, "Sensitivity analysis of stochastic models of bistable biochemical reactions," *Lecture Notes in Computer Science*, vol. 5016, pp. 1–20, 2008.

Vita

Hongxuan Zhang received the Bachelor's degree in Electronics & Information Technology from Tianjin University, Tianjin, China, in 2002. During his undergraduate study, he won the first award in China Undergraduate Mathematical Contest in Modeling (CUMCM) in 2000. In 2002, he enrolled in a joint program between the Institute for Infocomm Research and the National University of Singapore, to conduct research in computational electromagnetics. He joined the Department of Electrical & Computer Engineering (ECE) at Johns Hopkins University in 2004, where his initial research was focused on medical image analysis and computational anatomy. He also served as a teaching assistant in the ECE department. Since August 2007, he has been working on computational systems biology and has published two journal papers in this field. Between 2007 – 2009, he worked as a scientist or statistician for several companies in the health care industry.