A comparison of approximation techniques for variance-based sensitivity analysis of biochemical reaction systems

Hong-Xuan Zhang[1] **and John Goutsias**[*1]

[1] Whitaker Biomedical Engineering Institute, The Johns Hopkins University, Baltimore, MD 21218, USA

[*] Corresponding author

Email addresses:

HXZ: hxzhang@jhu.edu

JG: goutsias@jhu.edu

BMC Bioinformatics 2010, 11:246

Abstract

Background: Sensitivity analysis is an indispensable tool for the analysis of complex systems. In a recent paper, we have introduced a thermodynamically consistent variance-based sensitivity analysis approach for studying the robustness and fragility properties of biochemical reaction systems under uncertainty in the standard chemical potentials of the activated complexes of the reactions and the standard chemical potentials of the molecular species. In that approach, key sensitivity indices were estimated by Monte Carlo sampling, which is computationally very demanding and impractical for large biochemical reaction systems. Computationally efficient algorithms are needed to make variance-based sensitivity analysis applicable to realistic cellular networks, modeled by biochemical reaction systems that consist of a large number of reactions and molecular species.

Results: We present four techniques, derivative approximation (DA), polynomial approximation (PA), Gauss-Hermite integration (GHI), and orthonormal Hermite approximation (OHA), for *analytically* approximating the variance-based sensitivity indices associated with a biochemical reaction system. By using a well-known model of the mitogen-activated protein kinase signaling cascade as a case study, we numerically compare the approximation quality of these techniques against traditional Monte Carlo sampling. Our results indicate that, although DA is computationally the most attractive technique, special care should be exercised when using it for sensitivity analysis, since it may only be accurate at low levels of uncertainty. On the other hand, PA, GHI, and OHA are computationally more demanding than DA but can work well at high levels of uncertainty. GHI results in a slightly better accuracy than PA, but it is more difficult to implement. OHA produces the most accurate approximation results and can be implemented in a straightforward manner. It turns out that the computational cost of the four approximation techniques considered in this paper is orders of magnitude smaller than traditional Monte Carlo estimation. Software, coded in MATLAB®, which implements all sensitivity analysis techniques discussed in this paper, is available free of charge.

Conclusions: Estimating variance-based sensitivity indices of a large biochemical reaction system is a computationally challenging task that can only be addressed via approximations. Among the methods presented in this paper, a technique based on orthonormal Hermite polynomials seems to be an acceptable candidate for the job, producing very good approximation results for a wide range of uncertainty levels in a fraction of the time required by traditional Monte Carlo sampling.

Background

Sensitivity analysis is an indispensable tool for the analysis of complex systems [1, 2]. It is routinely used to investigate how uncertainty in input variables affects uncertainty in system response and to quantify the relative importance of the input variables in influencing the response. In addition to many other areas of science and engineering, sensitivity analysis is used in systems biology to investigate the robustness and fragility properties of cellular systems, such as signaling, gene regulation, and metabolic networks [3–11], as well as in systems pharmacology [12], for designing novel pharmacological intervention strategies and for understanding drug action [13, 14].

To study the sensitivity properties of a biochemical reaction system, such as a signaling network, we must construct a mathematical model that relates uncertainty in key biochemical factors of interest to a biologically relevant system response, and develop techniques for determining how factor uncertainty affects the system response. Since biochemical reaction systems are subject to physical laws, an important requirement is that sensitivity analysis must satisfy important thermodynamic constraints, such as the principle of detailed balance [15]. Bearing these in mind, we have proposed in [16] a probabilistic sensitivity analysis approach for biochemical reaction systems that uses the standard chemical potentials of the activated complexes of the underlying reactions and 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 [2, 17–19], 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 [2, 16, 18, 20], 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 [21]. 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 burden 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 paper, we present four such techniques and apply them to a well-known biochemical reaction model of the mitogen-activated protein kinase (MAPK) signalling cascade. The first technique 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 [2, 18, 19, 22] by including second-order derivative terms. The other approximation techniques are based on the high-dimensional model representation (HDMR) schemes developed by H. Rabitz and his coworkers [23–25]. We use analytical derivations, provided in the Additional file 1 accompanying this paper, 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. We have coded the sensitivity analysis techniques discussed in this paper using MATLAB®. Interested readers can request a copy of the software, and the entire set of data obtained with this software, by contacting the corresponding author.

We should mention here that, in systems biology, the most commonly used sensitivity analysis techniques are based on derivatives of molecular concentrations or other system responses, known as control coefficients [3]. These differential methods are based on a Taylor series approximation of the response function and, as such, are subject to several drawbacks that must be carefully considered before applying them to problems of systems biology. For example, derivative-based sensitivity indices assess the sensitivity properties of a biochemical reaction system around a set of reference input values. Their performance usually depends on the particular choice of these values, due to the nonlinear nature of the response function. For the results to be relevant, the reference values must be the true values, which are usually not known in practice. As a consequence, derivative-based sensitivity analysis techniques are limited by the quality of the underlying Taylor series approximation. Moreover, and due to our difficulty in accurately evaluating high-order derivatives, differential sensitivity analysis techniques are usually limited in practice to assessing the effect of one input factor on the system response, by keeping all other factors fixed to their reference values. This is usually not adequate, since we are most often interested in the effects of multiple biochemical factors on the system response. Finally, traditional differential analysis cannot cope with probabilistic uncertainty in biochemical factor values, unless it is combined with variance-based sensitivity analysis (as it is done by the first approximation technique considered in this paper). It turns out

that variance-based sensitivity analysis does not depend on the additivity or linearity of the system model and can be naturally used to quantify the simultaneous effect of probabilistic biochemical factor uncertainty on the system response [2, 18]. For this reason, it provides a very attractive and powerful approach for sensitivity analysis of biochemical reaction systems.

We should finally mention that a number of alternative approximation techniques for variance-based sensitivity analysis have been proposed in the literature [26–29]. In these techniques, the original response function is approximated by a surrogate function and the sensitivity indices are then estimated by Monte Carlo sampling based on that function. Reduction in computations is achieved by the fact that the time required for computing the system response at each Monte Carlo iteration using the surrogate function is much smaller than computing the response using the original function (whose evaluation requires solving a system of ordinary differential equations). However, the computations associated with these techniques are still substantial, since they must employ a large number of samples to sufficiently reduce the Monte Carlo estimation error. By contrast, the techniques discussed in this paper are based on surrogate functions that lead to analytical formulas for the sensitivity indices, thus avoiding Monte Carlo estimation. As a matter of fact, the computational cost for calculating the variance-based sensitivity indices using the techniques discussed in this paper is mainly associated with the problem of estimating the underlying parameters of the surrogate function used, which leads to appreciable computational savings over the techniques proposed in [26–29].

Methods

Biochemical reaction systems

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

$$\sum_{n=1}^{N} \nu_{nm} X_n \underset{\kappa_{2m}}{\overset{\kappa_{2m-1}}{\rightleftarrows}} \sum_{n=1}^{N} \nu'_{nm} X_n,\ m = 1, 2, \ldots, M,$$

where $\kappa_{2m-1}, \kappa_{2m} \geq 0$ are the normalized rate constants of the forward and reverse reactions (measured in s^{-1}) and $\nu_{nm}, \nu'_{nm} \geq 0$ are the stoichiometry coefficients of the reactants and products. We assume that the system consists of N molecular species $X_1, X_2, \ldots, X_N$, with concentrations (measured in molecules/cell) at time $t \geq 0$ given by $q_1(t)$, $q_2(t)$, …, $q_N(t)$, respectively. We characterize the dynamic evolution of

molecular concentrations by the following chemical kinetic equations:

$$\frac{dq_n(t)}{dt} = \sum_{m=1}^{M} s_{nm}\rho_m(t), \quad t \geq 0, \;\; n = 1, 2, \ldots, N, \tag{1}$$

where $s_{nm} := \nu'_{nm} - \nu_{nm}$ is the stoichiometry coefficient of the n^{th} molecular species associated with the m^{th} reaction and

$$\rho_m(t) := \kappa_{2m-1} \prod_{i=1}^{N} [q_i(t)]^{\nu_{im}} - \kappa_{2m} \prod_{i=1}^{N} [q_i(t)]^{\nu'_{im}} \tag{2}$$

is the flux of the m^{th} reaction at time t.

The sensitivity analysis approach we consider here is based on quantifying the influence of a reaction or molecular species on an appropriately chosen response characteristic R of a biochemical reaction system. We employ a well-known model of the MAPK signaling cascade (see Figure 1 and Additional file 2 for details on this model) and consider three response characteristics with established biological significance, namely the *duration* D, *integrated response* I, and *strength* S of the doubly phosphorylated extracellular signal-regulated kinase (ERK-PP), defined by

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

where $q(t)$ is the concentration profile of ERK-PP and t_0 is the time at which $q(t)$ converges to zero. If convergence to zero does not occur within the observation time interval $[0, t_{\max}]$, then we set $t_0 = t_{\max}$. We choose to work with the duration, integrated response, and strength of ERK-PP activity, since it has been experimentally observed that differences in duration and strength may cause distinct biological outcomes, such as cell differentiation, proliferation, and apoptosis [30–34], whereas, the integrated response directly correlates with DNA synthesis [35, 36]. We take the system response R to be the logarithm of the duration, integrated response, or strength; that is, we take R to be $\ln D$, $\ln I$, or $\ln S$. This reduces the effect of outliers and increases the efficiency of numerically evaluating the indices associated with variance-based sensitivity analysis [16].

Variance-based sensitivity analysis

We employ the variance-based sensitivity analysis approach for biochemical reaction systems we recently introduced in [16]. This method is based on a biophysically-derived probabilistic model for the rate constants of a biochemical reaction system. According to this model, we treat the rate constants κ_{2m-1} and κ_{2m} as random variables K_{2m-1} and K_{2m}, given by the Eyring-Polanyi equations [37]

$$\begin{aligned} K_{2m-1} &= \frac{k_B T}{h} \frac{C_m^{\ddagger}}{\prod_{n=1}^{N} C_n^{\nu_{nm}}} \\ K_{2m} &= \frac{k_B T}{h} \frac{C_m^{\ddagger}}{\prod_{n=1}^{N} C_n^{\nu'_{nm}}} , \end{aligned} \tag{3}$$

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

$$\begin{aligned} C_m^{\ddagger} &:= \exp\left\{-\frac{M_m^{\ddagger 0}}{k_B T}\right\} \\ C_n &:= \exp\left\{-\frac{M_n^{0}}{k_B T}\right\}, \end{aligned} \tag{4}$$

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

$$\begin{aligned} M_m^{\ddagger 0} &= \mu_m^{\ddagger 0} + k_B T Y_m^{\ddagger} \\ M_n^{0} &= \mu_n^{0} + k_B T Y_n. \end{aligned} \tag{5}$$

In (5), $\mu_m^{\ddagger 0}$ and μ_n^0 are the nominal standard chemical potential values associated with the m^{th} reaction and the n^{th} molecular species, 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 variations in the standard chemical potentials about their nominal values caused by unpredictable biological variability and uncertainty regarding their exact values. Similarly to [16], we assume that the random variables $Y_m^{\ddagger}, m = 1, 2, \ldots, M$, and $Y_n, n = 1, 2, \ldots, N$, are statistically independent.

Our variance-based sensitivity analysis technique assesses how uncertainty in the rate constants of a biochemical reaction system affect the system response. As a consequence of (3), (4), and (5), we have that

$$K_{2m-1} \;=\; \kappa_{2m-1}\exp\{-Y_m^{\ddagger}\}\exp\left\{\sum_{n=1}^{N}\nu_{nm}Y_n\right\}$$

$$K_{2m} \;=\; \kappa_{2m}\exp\{-Y_m^{\ddagger}\}\exp\left\{\sum_{n=1}^{N}\nu'_{nm}Y_n\right\}, \tag{6}$$

where

$$\kappa_{2m-1} \;=\; \frac{k_BT}{h}\frac{c_m^{\ddagger}}{\prod_{n=1}^{N}c_n^{\nu_{nm}}}$$

$$\kappa_{2m} \;=\; \frac{k_BT}{h}\frac{c_m^{\ddagger}}{\prod_{n=1}^{N}c_n^{\nu'_{nm}}}$$

are the nominal values of the rate constants, with

$$c_m^{\ddagger} \;:=\; \exp\left\{-\frac{\mu_m^{\ddagger 0}}{k_BT}\right\}$$

$$c_n \;:=\; \exp\left\{-\frac{\mu_n^{0}}{k_BT}\right\}.$$

Equation (6) suggests that uncertainty in the forward and reverse reaction rates occurs due to uncertainty in the standard chemical potentials of the activated complexes associated with the reactions and the standard chemical potentials of the reactants.

As a consequence of the previous model, we investigate the sensitivity properties of a biochemical reaction system due to the uncertainty in the standard chemical potentials. To simplify notation, we use $\boldsymbol{W} = \{W_1, W_2, \ldots, W_J\}$ to denote the random variables $Y^{\ddagger}$ and Y. We consider two cases, namely $J = M$ and $W_j = Y_j^{\ddagger}$, for $j = 1, 2, \ldots, M$, as well as $J = N$ and $W_j = Y_j$, $j = 1, 2, \ldots, N$. In the first case, the standard chemical potentials of the molecular species are assumed to be fixed, whereas, the standard chemical potentials of the activated complexes are perturbed randomly. Our objective is to investigate the importance of reactions in influencing the system response and, for this reason, we refer to this case as *reaction-oriented sensitivity analysis* (ROSA) [16]. In the second case, the standard chemical potentials of the activated complexes are assumed to be fixed, whereas, the standard chemical potentials of the molecular species are perturbed randomly. In this case, our objective is to investigate the importance of

molecular species in influencing the system response. For this reason, we refer to this case as *species-oriented sensitivity analysis* (SOSA) [16].

Given the response R of a biochemical reaction system with random factors $\boldsymbol{W}$, its total variance $V_{\text{tot}} := \text{Var}[R(\boldsymbol{W})]$ satisfies [17, 38, 39]:

$$V_{\text{tot}} = \sum_{j=1}^{J} V_j + \sum_{j=1}^{J-1} \sum_{j'=j+1}^{J} V_{jj'} + \cdots + V_{12\cdots J}, \tag{7}$$

where

$$\begin{aligned} V_j &:= \text{Var}[\text{E}[R(\boldsymbol{W}) \mid W_j]] \\ V_{jj'} &:= \text{Var}[\text{E}[R(\boldsymbol{W}) \mid W_j, W_{j'}]] - V_j - V_{j'}, \end{aligned} \tag{8}$$

with similar expressions for the remaining terms. If the biochemical factors $\boldsymbol{W}$ are statistically independent (which we assume here to be true), then each term on the right-hand-side of (7) is nonnegative. This equation provides a decomposition of the total system response variance V_{tot} into individual terms $V_1, V_2, \ldots, V_{12}, \ldots$. It turns out that V_j quantifies the average reduction in total response variance, obtained by keeping the j^{th} biochemical factor fixed. As a consequence, we use V_j to measure the *singular* influence of the j^{th} biochemical factor W_j on the system response. Moreover, the term $V_{jj'}$ quantifies the average reduction in the total response variance due to jointly fixing the two biochemical factors W_j and $W_{j'}$, not accounted for by summing the average reductions obtained by separately fixing these factors. Therefore, we use $V_{jj'}$ to measure the *joint* influence of the biochemical factors W_j and $W_{j'}$ on the system response. Finally, higher-order terms in (7) quantify the joint influence of three or more biochemical factors on 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 (7). Although these terms are usually negligible at low to moderate levels of biochemical factor uncertainty, they may take substantial values at high levels [16]. 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_{tot}, we base our sensitivity analysis on its second-order portion V, given by

$$V = \sum_{j=1}^{J} V_j + \sum_{j=1}^{J-1} \sum_{j'=j+1}^{J} V_{jj'}. \tag{9}$$

By using the probabilistic model given by (6) and the variance decomposition scheme in (9), we can develop a powerful (second-order) methodology for sensitivity analysis of biochemical reaction systems, similar to the one discussed in [16] that was based on the total response variance V_{tot}. The method requires evaluation of two indices, namely the (second-order) *single-effect sensitivity index* (SESI) σ_j, defined by

$$\sigma_j := \frac{V_j}{V}, \tag{10}$$

and the (second-order) *joint-effect sensitivity index* (JESI) η_j, defined by

$$\eta_j := \frac{U_j}{V}, \tag{11}$$

where

$$U_j := \sum_{j=1, j' \neq j}^{J} V_{jj'}. \tag{12}$$

Clearly, σ_j quantifies the fractional singular contribution of the j^{th} biochemical factor to the second-order portion V of the total response variance, whereas, η_j quantifies the fractional contribution of the j^{th} biochemical factor to V jointly with another factor. It turns out that, if $\sigma_j = \eta_j = 0$, 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 $\sigma_j > 0$ and $\eta_j = 0$, then we can conclude that factor j influences the system response singularly but not jointly with another factor. Moreover, if $\sigma_j = 0$ and $\eta_j > 0$, we can conclude that factor j does not influence the system response singularly but it does so jointly with some other factor, whereas, if $\sigma_j > 0$ and $\eta_j > 0$, 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 σ_j and η_j are sufficiently larger than zero.

Unfortunately, we cannot evaluate the exact values of the sensitivity indices σ_j and η_j. For this reason, we must resort to approximations. In this paper, we consider the possibility of employing one of five methods to accomplish this goal. We discuss these methods next and refer the reader to [16] and the accompanying Additional file 1 for details pertaining to their development and numerical implementation.

Monte Carlo estimation

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 [16] (see also [2, 20]). This approach can be used to provide estimates $\widehat{\sigma}_j$ and $\widehat{\eta}_j$ 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 (1) and (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 $\widehat{\sigma}_j$ and $\widehat{\eta}_j$ 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.

Derivative approximation

A method for deriving approximations $\widehat{\sigma}_j$ and $\widehat{\eta}_j$ of the sensitivity indices σ_j and η_j is to replace the response function $R(\boldsymbol{w})$ by its second-order Taylor series approximation $\widehat{R}(\boldsymbol{w})$ about $\boldsymbol{w} = \mathbf{0}$, given by

$$\widehat{R}(\boldsymbol{w}) = R(\mathbf{0}) + \sum_{j=1}^{J} d_j w_j + \frac{1}{2} \sum_{j=1}^{J} \sum_{j'=1}^{J} d_{jj'} w_j w_{j'}, \tag{13}$$

where

$$d_j := \frac{\partial R(\mathbf{0})}{\partial w_j} \qquad \text{and} \qquad d_{jj'} := \frac{\partial^2 R(\mathbf{0})}{\partial w_j \partial w_{j'}}$$

are the first- and second-order partial derivatives of R at $\boldsymbol{w} = \mathbf{0}$, and set

$$\widehat{\sigma}_j := \frac{\widehat{V}_j}{\widehat{V}}, \qquad \widehat{\eta}_j := \frac{\widehat{U}_j}{\widehat{V}}, \tag{14}$$

where

$$\begin{aligned}
\widehat{V} &:= \sum_{j=1}^{J} \widehat{V}_j \; + \; \sum_{j=1}^{J-1} \sum_{j'=j+1}^{J} \widehat{V}_{jj'} \\
\widehat{U}_j &:= \sum_{j'=1, j' \neq j}^{J} \widehat{V}_{jj'} \\
\widehat{V}_j &= \mathrm{Var}[\mathrm{E}[\widehat{R}(\boldsymbol{W}) \mid W_j]] \\
\widehat{V}_{jj'} &= \mathrm{Var}[\mathrm{E}[\widehat{R}(\boldsymbol{W}) \mid W_j, W_{j'}]] - \widehat{V}_j - \widehat{V}_{j'}.
\end{aligned} \tag{15}$$

Equation (13) and the statistical independence and zero-mean Gaussianity of the biochemical factors W_j imply that

$$\widehat{V}_j = \lambda_j^2 d_j^2 + \frac{1}{2} \lambda_j^4 d_{jj}^2 \qquad \text{and} \qquad \widehat{V}_{jj'} = \lambda_j^2 \lambda_{j'}^2 d_{jj'}^2, \tag{16}$$

where λ_j is the standard deviation of W_j, for $j = 1, 2, \ldots, J$. As a consequence, we obtain an analytical expression for the sensitivity indices $\widehat{\sigma}_j$ and $\widehat{\eta}_j$, which requires evaluation of the first- and second-order partial derivatives of the response function $R(\boldsymbol{w})$, with respect to the biochemical factors, at $\boldsymbol{w} = \mathbf{0}$.

Although many techniques have been developed to compute response derivatives [40], for reasons we explain in Additional file 1, we choose to approximate the partial derivatives by symmetric finite differences. We refer to $\widehat{\sigma}_j$ and $\widehat{\eta}_j$, given by (14), (15), and (16), as the (second-order) SESI's and JESI's obtained by *Derivative Approximation* (DA). The resulting method requires $2J(J+1)+1$ system integrations, which is quadratic in terms of the number J of the biochemical factors and is much smaller than the number $2L(J+1)$ of system integrations required by MC, since $J \ll L$.

Polynomial approximation

Another way to approximate the sensitivity indices σ_j and η_j is to replace the response function $R(\boldsymbol{w})$ by

$$\begin{aligned}\widehat{R}(\boldsymbol{w}) = R(\mathbf{0}) & \\ &+\sum_{j=1}^{J}(\alpha_{j,1}w_j+\alpha_{j,2}w_j^2) \\ &+\sum_{j=1}^{J-1}\sum_{j'=j+1}^{J}\alpha_{jj',1}w_jw_{j'}+\sum_{j=1}^{J-1}\sum_{j'=j+1}^{J}\alpha_{jj',2}w_j^2w_{j'} \\ &+\sum_{j=1}^{J-1}\sum_{j'=j+1}^{J}\alpha_{jj',3}w_jw_{j'}^2+\sum_{j=1}^{J-1}\sum_{j'=j+1}^{J}\alpha_{jj',4}w_j^2w_{j'}^2,\end{aligned} \tag{17}$$

where the α's are parameters whose values must be appropriately determined so that $\widehat{R}(\boldsymbol{w})$ sufficiently approximates the response function $R(\boldsymbol{w})$ in an appropriately chosen neighborhood around $\mathbf{0}$. Note that $\widehat{R}(\boldsymbol{w})$ provides a polynomial approximation of the response function $R(\boldsymbol{w})$. If $\widehat{R}(\boldsymbol{w})$ is sufficiently close to $R(\boldsymbol{w})$ in a neighborhood around $\mathbf{0}$, then the parameters α coincide with the partial derivatives $\partial^{\kappa_1+\kappa_2}R(\mathbf{0})/\partial w_j^{\kappa_1}\partial w_{j'}^{\kappa_2}$, $1 \le \kappa_1, \kappa_2 \le 2$, of R at $\boldsymbol{w}=\mathbf{0}$.

By using (17) and the statistical independence and zero-mean Gaussianity of the biochemical factors W_j, we can show that, in this case, $\widehat{\sigma}_j$ and $\widehat{\eta}_j$ are given by (14) and (15), with

$$\begin{aligned}\widehat{V}_j =& \lambda_j^2\alpha_{j,1}^2+2\lambda_j^4\alpha_{j,2}^2 \\ &+2\lambda_j^2\alpha_{j,1}\left(\sum_{m=1}^{j-1}\lambda_m^2\alpha_{mj,2}+\sum_{m=j+1}^{J}\lambda_m^2\alpha_{jm,3}\right) \\ &+\lambda_j^2\left(\sum_{m=1}^{j-1}\lambda_m^2\alpha_{mj,2}+\sum_{m=j+1}^{J}\lambda_m^2\alpha_{jm,3}\right)^2\end{aligned}$$

$$+\ 4\lambda_j^4\alpha_{j,2}\left(\sum_{m=1}^{j-1}\lambda_m^2\alpha_{mj,4}+\sum_{m=j+1}^{J}\lambda_m^2\alpha_{jm,4}\right)$$
$$+\ 2\lambda_j^4\left(\sum_{m=1}^{j-1}\lambda_m^2\alpha_{mj,4}+\sum_{m=j+1}^{J}\lambda_m^2\alpha_{jm,4}\right)^2$$

$$\widehat{V}_{jj'}=\lambda_j^2\lambda_{j'}^2\alpha_{jj',1}^2+2\lambda_j^4\lambda_{j'}^2\alpha_{jj',2}^2$$
$$+2\lambda_j^2\lambda_{j'}^4\alpha_{jj',3}^2+4\lambda_j^4\lambda_{j'}^4\alpha_{jj',4}^2. \tag{18}$$

As a consequence, we obtain again an analytical expression for the sensitivity indices $\widehat{\sigma}_j$ and $\widehat{\eta}_j$, which requires evaluation of the α parameters. This can be done by the polynomial regression approach we discuss in Additional file 1. We refer to $\widehat{\sigma}_j$ and $\widehat{\eta}_j$, given by (14), (15), and (18), as the (second-order) SESI's and JESI's obtained by *Polynomial Approximation* (PA). The resulting method is based on the approach proposed in [41] and 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 DA, since $S>2$.

Gauss-Hermite integration

We can obtain a more accurate approximation $\widehat{R}(\boldsymbol{w})$ of the response function $R(\boldsymbol{w})$ than the one given by (13) if we truncate the Taylor series expansion of $R(\boldsymbol{w})$ about $\boldsymbol{w}=\mathbf{0}$ by removing all terms that involve partial derivatives with respect to more than two factors [note that the approximation given by (13) is obtained from the Taylor series expansion by truncating all terms that involve partial derivatives of order greater than two]. In this case, we can show that

$$\widehat{R}(\boldsymbol{w})=\psi_0\ -\ (J-2)\sum_{j=1}^{J}\psi_j(w_j)$$
$$+\sum_{j=1}^{J-1}\sum_{j'=j+1}^{J}\psi_{jj'}(w_j,w_{j'}),$$

where

$$\psi_0 := \frac{(J-1)(J-2)}{2} R(0,0,\ldots,0)$$

$$\psi_j(w_j) := R(0,\ldots,w_j,\ldots,0)$$

$$\psi_{jj'}(w_j,w_{j'}) := R(0,\ldots,w_j,\ldots,w_{j'},\ldots,0), \tag{19}$$

as we explain in Additional file 1. The approximations $\widehat{\sigma}_j$ and $\widehat{\eta}_j$ are now given by (14) and (15), with

$$\begin{aligned} \widehat{V}_j &= \mathrm{E}[e_j^2(W_j)] - e_0^2 \\ \widehat{V}_{jj'} &= \mathrm{E}[e_{jj'}^2(W_j,W_{j'})] - \widehat{V}_j - \widehat{V}_{j'} - e_0^2, \end{aligned} \tag{20}$$

where e_0, e_j, and $e_{jj'}$ are given by

$$\begin{aligned} e_0 &= \psi_0 - (J-2)\sum_{m=1}^{J} \mathrm{E}[\psi_m(W_m)] \\ &\quad + \sum_{m=1}^{J-1}\sum_{m'=m+1}^{J} \mathrm{E}[\psi_{mm'}(W_m,W_{m'})] \end{aligned}$$

$$\begin{aligned} e_j(w_j) &= \psi_0 \\ &-(J-2)\sum_{m=1}^{J} \mathrm{E}[\psi_m(W_m) \mid W_j = w_j] \\ &+ \sum_{m=1}^{J-1}\sum_{m'=m+1}^{J} \mathrm{E}[\psi_{mm'}(W_m,W_{m'}) \mid W_j = w_j] \end{aligned}$$

$$\begin{aligned} e_{jj'}(w_j,w_{j'}) &= \psi_0 \\ &-(J-2)\sum_{m=1}^{J} \mathrm{E}[\psi_m(W_m) \mid W_j = w_j, W_{j'} = w_{j'}] \\ &+ \sum_{m=1}^{J-1}\sum_{m'=m+1}^{J} \mathrm{E}[\psi_{mm'}(W_m,W_{m'}) \mid \\ &\qquad\qquad W_j = w_j, W_{j'} = w_{j'}]. \end{aligned} \tag{21}$$

Note that evaluation of $\widehat{V}_j$ and $\widehat{V}_{jj'}$ requires only one- and two-dimensional integrations, which can be numerically done by a standard Gauss-Hermite integration approach. For this reason, we refer to $\widehat{\sigma}_j$ and $\widehat{\eta}_j$, given by (14), (15), (19), (20), and (21), as the (second-order) SESI's and JESI's obtained by *Gauss*

Hermite Integration (GHI). The resulting method is based on the approach proposed in [42, 43] and requires $2J(J-1)\lfloor Q/2\rfloor^2 + 2J\lfloor Q/2\rfloor + 1$ system integrations, which is quadratic both in terms of the number J of biochemical factors and the order Q of Gauss-Hermite integration used. Note that, if the number S of the samples per factor used in the regression associated with the PA is even, and $Q = S$ or $Q = S + 1$, then GHI requires the same number of system integrations as PA.

Orthonormal Hermite approximation

The last method we consider for approximating the sensitivity indices σ_j and η_j is based on replacing the response function $R(\boldsymbol{w})$ by

$$\begin{aligned}\widehat{R}(\boldsymbol{w}) = \widehat{\rho}_0 &\\ &+\sum_{j=1}^{J}\left[\alpha_{j,1}\frac{w_j}{\lambda_j} + \frac{\alpha_{j,2}}{\sqrt{2}}\left(\frac{w_j^2}{\lambda_j^2} - 1\right)\right]\\ &+\sum_{j=1}^{J-1}\sum_{j'=j+1}^{J}\alpha_{jj',1}\frac{w_j w_{j'}}{\lambda_j\lambda_{j'}}\\ &+\sum_{j=1}^{J-1}\sum_{j'=j+1}^{J}\frac{\alpha_{jj',2}}{\sqrt{2}}\left(\frac{w_j^2}{\lambda_j^2} - 1\right)\frac{w_{j'}}{\lambda_{j'}}\\ &+\sum_{j=1}^{J-1}\sum_{j'=j+1}^{J}\frac{\alpha_{jj',3}}{\sqrt{2}}\frac{w_j}{\lambda_j}\left(\frac{w_{j'}^2}{\lambda_{j'}^2} - 1\right)\\ &+\sum_{j=1}^{J-1}\sum_{j'=j+1}^{J}\frac{\alpha_{jj',4}}{2}\left(\frac{w_j^2}{\lambda_j^2} - 1\right)\left(\frac{w_{j'}^2}{\lambda_{j'}^2} - 1\right),\end{aligned} \tag{22}$$

where the α's and $\widehat{\rho}_0$ are parameters whose values must be appropriately determined so that $\widehat{R}(\boldsymbol{w})$ sufficiently approximates the response function $R(\boldsymbol{w})$ over the entire space of biochemical factor values. Note that $\widehat{R}(\boldsymbol{w})$ provides a polynomial approximation of the response function $R(\boldsymbol{w})$, similar to the one given by (17). However, the polynomials used in the approximation given by (22) are orthonormal Hermite polynomials, as opposed to the polynomials used in the approximation given by (17), which are standard second- and fourth-order polynomials. Note also that the approximation given by (22) is "global," in the sense that it is based on approximating the system response function $R(\boldsymbol{w})$ over the entire factor space, whereas, the approximation given by (17) is "local," in the sense that it approximates the system response function $R(\boldsymbol{w})$ in a neighborhood around $\boldsymbol{w} = \boldsymbol{0}$.

By using (22), the orthonormality of the Hermite polynomials, and the statistical independence and zero-mean Gaussianity of the biochemical factors W_j, we can show that $\widehat{\sigma}_j$ and $\widehat{\eta}_j$ are given by (14) and (15), with

$$\begin{aligned} \widehat{V}_j &= \alpha_{j,1}^2 + \alpha_{j,2}^2 \\ \widehat{V}_{jj'} &= \alpha_{jj',1}^2 + \alpha_{jj',2}^2 + \alpha_{jj',3}^2 + \alpha_{jj',4}^2. \end{aligned} \tag{23}$$

As a consequence, we obtain again an analytical expression for the sensitivity indices $\widehat{\sigma}_j$ and $\widehat{\eta}_j$, which requires evaluation of the α parameters. This can be done by polynomial regression based on the Monte Carlo Latin hypercube sampling approach we discuss in Additional file 1. We refer to $\widehat{\sigma}_j$ and $\widehat{\eta}_j$, given by (14), (15), and (23), as the (second-order) SESI's and JESI's obtained by *Orthonormal Hermite Approximation* (OHA). The resulting method is based on the approach suggested in [44–47] and requires L system integrations, where L is the number of regression points obtained by Latin hypercube sampling. We here take the number of regression points used to be the same as the number of Latin hypercube samples employed by MC, although these two numbers can be different in general. As a consequence, the number of system integrations performed by OHA 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.

Results

We now employ the previously discussed techniques to estimate the variance-based sensitivity indices σ_j and η_j 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 Figure 1 (see Additional file 2 for more details on this model). As we have explained in the previous section, we consider two cases: ROSA and SOSA. 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 perturbation levels in biochemical factor values. By following our previous work in [16], we perform sensitivity analysis with $\lambda^\ddagger, \lambda = 0.1, 0.2, 0.3, 0.4$. Finally, we employ $L = 6{,}000$ Latin hypercube samples in MC and OHA, $S = 4$ regression samples per factor in PA, and a Gauss-Hermite integration of order $Q = 5$ in GHI.

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 $\mathbf{0}$). Note also that, as a consequence of equation (6), 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 $100[\exp\{-\lambda_m^{\ddagger}\} - 1, \exp\{\lambda_m^{\ddagger}\} - 1]\%$. 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 $\lambda^{\ddagger} = 0.4$.

In Table 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 2, 3, and 4, whereas, we list the SOSA results in Tables S-3.1, S-3.2, and S-3.3 of Additional file 3. 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 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 our previous work in [16], we classify reactions and molecular species into one of four categories of interest: singularly influential, jointly influential, singularly/jointly influential, and noninfluential. We do this by comparing their SESI and JESI values to a 10% threshold. Bold reaction numbers indicate SESI or JESI values, obtained by MC, that are above the 10% threshold. Note that 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 noninfluential if both the SESI and JESI values are smaller than 10%.

In the remaining of this section, we discuss the ROSA results separately for each response characteristic. A similar discussion applies for the SOSA results presented in Additional file 3.

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 perturbations; see Table 2. Moreover, the estimated SESI and JESI values indicate that the duration is primarily affected by reactions 4, 6, and 13 (refer to Figure 1 and Additional file 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, at least within the range of the applied perturbations. Note that a multivariate response function is called additive if it can be decomposed into a sum of one-dimensional functions of one variable. Additive response functions do not produce high-order (≥ 2) joint effects and result in zero JESI values [2]. Although a linear response function is additive, the inverse is not necessarily true. 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 2 (and Table S-3.1 in Additional file 3), 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 perturbation levels. Moreover, all methods can be used to correctly classify reactions 4, 6, and 13 as being singularly influential.

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 3. These values indicate that the integrated response is primary influenced by reactions 4 and 6 (refer to Figure 1 and Additional file 2 for identifying these reactions). For small to moderate perturbations (i.e., for $\lambda^{\ddagger} = 0.1, 0.2$), reactions 4 and 6 influence the integrated response only singularly. However, for large perturbations (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 the integrated response only singularly.

It is clear from the results depicted in Table 3 (and Table S-3.2 in Additional file 3) that all approximation techniques work relatively well for small to moderate perturbation 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 perturbations (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, deeming the use of DA inappropriate. 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. 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 3 (and Table S-3.2 in Additional file 3), it is clear that, for large perturbations, 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 perturbation 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.

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. 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 perturbation 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 noninfluential, 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 (and Table S-3.3 in Additional file 3) 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. For example, using 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. Once more, OHA consistently provides good results, which can be used to correctly classify the reactions at all perturbation levels.

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 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. 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 ("curse of dimensionality").

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 perturbations, 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 perturbations. 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 perturbations, the JESI values obtained by DA, associated with the integrated response and strength, are significantly different than the ones produced by MC.

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. In our simulations, we approximate the first- and second-order partial derivatives of the response function by using equations (S-1.35) and (S-1.36) in Additional file 1, with $\Delta = 0.1$. It has been pointed out in [1] 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 [40], 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 [1].

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 (see Additional file 1 for details). Errors are introduced by truncating the FD-HDMR and *locally* approximating the resulting response function by a fourth-order polynomial including only single biochemical factors and pairs of factors. As a consequence, PA may not be able to accurately estimate some SESI and JESI values under large perturbations, since the underlying truncation and polynomial approximation of the response function may not be sufficiently accurate over the range of factor values generated by such perturbations. Note also that errors can be introduced due to estimating the polynomial coefficients by regression, a situation that cannot be evaluated and controlled easily. As a matter of fact, and counter to intuition, we cannot necessarily increase accuracy of estimation by using more samples per biochemical factor, especially when dealing with polynomial regression [48, 49].

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 (see Additional file 1 for details). 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. Note that 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 can be well-approximated by polynomials [49]. 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 perturbations, since the underlying FD-HDMR truncation may not be sufficiently accurate over the range of factor values generated by such perturbations.

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 (see Additional file 1 for details). Here, the truncation of high-order ANOVA-HDMR components does not 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 perturbation levels at which the higher-order (≥ 3) terms in the variance decomposition scheme given by (7) are negligible, then the higher-order (≥ 3) terms in the ANOVA-HDMR decomposition of the response function will be negligible as well [see equation (S-1.30) in Additional file 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 perturbation levels at which the response variance is not appreciably influenced by high-order joint effects. Under very large perturbations, 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 (7), 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 paper.

Although we have also obtained simulation results for other biochemical reaction systems, due to lack of space, we have limited our presentation in this paper to the results obtained for the MAPK model depicted

in Figure 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 paper, 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 paper. 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 paper are based on the assumption that, for most biochemical reaction systems of interest, perturbations 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 perturbation levels used for sensitivity analysis. In general, the approximation methods discussed in this paper 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 (7), 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 [42, 43], which has been effectively used to calculate statistical moments of the responses of high-dimensional mechanical systems subject to randomly fluctuating loads. In this paper, we have reformulated this method to fit the framework of variance-based sensitivity analysis and have applied it to biochemical reaction systems. On the other hand, OHA is based on the methodology proposed in [25, 44, 45, 50] 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 [46, 47, 51], and has been recently employed in [52] for estimating variance-based sensitivity indices in order to learn the topology of a functional network of interactions from given data. To our knowledge, this is the first time that the four approximation

techniques presented in this paper are systematically compared to each other and used to study the sensitivity properties of biochemical reaction systems.

To conclude, we would like to stress the fact that the approximation techniques presented in this paper 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 [16]. 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 paper. Developing variance-based sensitivity analysis for correlated biochemical factors is a challenging problem that needs careful investigation [2, 53]. 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 distributions. 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 [49]. Moreover, if the biochemical factors follow gamma distributions, then we must replace the orthonormal Hermite polynomials in OHA by orthonormal Laguerre polynomials [47, 51].

Conclusions

In this paper, we discussed four methods that one can use to analytically approximate the second-order sensitivity indices associated with a previously introduced variance-based sensitivity analysis methodology for biochemical reaction systems. The need for developing such methods stems from an effort to remedy the large computational burden associated with Monte Carlo estimation. 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 for the MAPK signalling 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 approximating the first- and second-order ANOVA-HDMR component functions with orthonormal Hermite polynomials.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

JG performed the design of the study and drafted the manuscript. HXZ contributed significantly to the manuscript by coding all methods in Matlab, by acquiring data for the study, and by interpreting the results. Both authors read and approved the final manuscript.

Acknowledgements

The authors acknowledge the National Science Foundation (NSF) for support of this research. Special thanks to Garrett Jenkinson for providing the C++ code used to develop a fast Matlab implementation of the MAPK signaling cascade model.

References

1. Varma A, Morbidelli M, Wu H: *Parametric Sensitivity in Chemical Systems*. Cambridge, UK: Cambridge University Press 1999.
2. Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, Saisana M, Tarantola S: *Global Sensitivity Analysis: The Primer*. Chichester, UK: John Wiley 2008.
3. Heinrich R, Schuster S: *The Regulation of Cellular Systems*. New York: Chapman & Hall 1996.
4. Cho KH, Shin SY, Kolch W, Wolkenhauer O: **Experimental design in systems biology, based on parameter sensitivity analysis using a Monte Carlo method: A case study for the TNFα-mediated NF-κB signal transduction pathway**. *Simulation* 2003, **79**:726–739.
5. Degenring D, Froemel C, Dikta G, Takors R: **Sensitivity analysis for the reduction of complex metabolism models**. *J. Process. Contr.* 2004, **14**:729–745.
6. Feng XJ, Hooshangi S, Chen D, Li G, Weiss R, Rabitz H: **Optimizing genetic circuits by global sensitivity analysis**. *Biophys. J.* 2004, **87**:2195–2202.
7. Leloup JC, Goldbeter A: **Modeling the mammalian circadian clock: Sensitivity analysis and multiplicity of oscillatory mechanisms**. *J. Theor. Biol.* 2004, **230**:541–562.
8. Stelling J, Gilles ED, Doyle FJ: **Robustness properties of circadian clock architectures**. *P. Natl. Acad. Sci. USA* 2004, **101**:13210–13215.
9. Hornberg JJ, Binder B, Bruggeman FJ, Schoeberl B, Heinrich R, Westerhoff HV: **Control of MAPK signalling: From complexity to what really matters**. *Oncogene* 2005, **24**:5533–5542.
10. Hu D, Yuan JM: **Time-dependent sensitivity analysis of biological networks: Coupled MAPK and PI3K signal transduction pathways**. *J. Chem. Phys.* 2006, **110**:5361–5370.
11. Mahdavi A, Davey RE, Bhola P, Yin T, Zandstra PW: **Sensitivity analysis of intracellular signaling pathway kinetics predicts targets for stem cell fate control**. *PLoS Comput. Biol.* 2007, **3**:1257–1267.
12. Berger SI, Iyengar R: **Network analyses in systems pharmacology**. *Bioinformatics* 2009, **25**:2466–2472.

13. Krewski D, Wang Y, Bartlett S, Krishnan K: **Uncertainty, variability, and sensitivity analysis in physiological pharmacokinetic models**. *J. Biopharm. Stat.* 1995, **5**:245–271.

14. Nestorov I: **Whole body pharmacokinetic models**. *Clin. Pharmacokinet.* 2003, **42**:883–908.

15. Ederer M, Gilles ED: **Thermodynamically feasible kinetic models of reaction networks**. *Biophys. J.* 2007, **92**:1846–1857.

16. Zhang HX, Dempsey WP, Goutsias J: **Probabilistic sensitivity analysis of biochemical reaction systems**. *J. Chem. Phys.* 2009, **131**:094101.

17. Sobol' IM: **Sensitivity estimates for nonlinear mathematical models**. *Math. Mod. Comput. Exp.* 1993, **1**:407–414.

18. Saltelli A, Tarantola S, Campolongo F, Ratto M: *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. Chichester, UK: John Wiley 2004.

19. Saltelli A, Ratto M, Tarantola S, Campolongo F: **Sensitivity analysis for chemical models**. *Chem. Rev.* 2005, **105**:2811–2827.

20. Saltelli A: **Making best use of model evaluations to compute sensitivity indices**. *Comput. Phys. Commun.* 2002, **145**:280–297.

21. Liu JS: *Monte Carlo Strategies in Scientific Computing*. New York: Springer 2001.

22. Helton JC, Davis FJ: **Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems**. *Reliab. Eng. Syst. Safe.* 2003, **81**:23–69.

23. Rabitz H, Alis ÖF, Shorter J, Shim K: **Efficient input-output model representations**. *Comput. Phys. Commun.* 1999, **117**:11–20.

24. Rabitz H, Alis ÖF: **General foundations of high-dimensional model representations**. *J. Math. Chem.* 1999, **25**:197–233.

25. Li G, Rosenthal C, Rabitz H: **High dimensional model representations**. *J. Phys. Chem. A* 2001, **105**:7765–7777.

26. Oakley JE, O'Hagan A: **Probabilistic sensitivity analysis of complex models: a Bayesian approach**. *J. Roy. Stat. Soc. B* 2004, **66**(3):751–769.

27. Ratto M, Pagano A, Young P: **State dependent parameter metamodelling and sensitivity analysis**. *Comput. Phys. Commun.* 2007, **177**:863–876.

28. Storlie CB, Helton JC: **Multiple predictor smoothing methods for sensitivity analysis: Description of techniques**. *Reliab. Eng. Syst. Safe.* 2008, **93**:28–54.

29. Storlie CB, Swiler LP, Helton JC, Sallaberry CJ: **Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models**. *Reliab. Eng. Syst. Safe.* 2009, **94**:1735–1763.

30. Marshall CJ: **Specificity of receptor tyrosine kinase signaling: Transient versus sustained extracellular signal-regulated kinase activation**. *Cell* 1995, **80**:179–185.

31. Murphy LO, Smith S, Chen RH, Fingar DC, Blenis J: **Molecular interpretation of ERK signal duration by immediate early gene products**. *Nat. Cell Biol.* 2002, **4**:556–564.

32. Murphy LO, MacKeigan JP, Blenis J: **A network of immediate early gene products propagates subtle differences in mitogen-activated protein kinase signal amplitude and duration**. *Mol. Cell Biol.* 2004, **24**:144–153.

33. Mayawala K, Gelmi CA, Edwards JS: **MAPK cascade possesses decoupled controllability of signal amplification and duration**. *Biophys. J.* 2004, **87**:L01–L02.

34. Ebisuya M, Kondoh K, Nishida E: **The duration, magnitude and compartmentalization of ERK MAP kinase activity: Mechanisms for providing signaling specificity**. *J. Cell Sci.* 2005, **118**:2997–3002.

35. Tombes RM, Auer KL, Mikkelsen R, Valerie K, Wymann MP, Marshall CJ, McMahon M, Dent P: **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**. *Biophys. J.* 1998, **330**:1451–1460.

36. Asthagiri AR, Reinhart CA, Horwitz AF, Lauffenburger DA: **The role of transient ERK2 signals in fibronectin- and insulin-mediated DNA synthesis**. *J. Cell Sci.* 2000, **113**:4499–4510.

37. Berry RS, Rice SA, Ross J: *Physical Chemistry*. New York: Oxford University Press, 2nd edition 2000.

38. Sobol' IM: **Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates**. *Math. Comput. Simulat.* 2001, **55**:271–280.

39. Sobol' IM: **Theorems and examples on high-dimensional model representation**. *Reliab. Eng. Syst. Safe.* 2003, **79**:187–193.

40. Cacuci DG: *Sensitivity and Uncertainty Analysis*, *Volume I. Theory*. Boca Raton: Chapman & Hall/CRC 2003.

41. Chen W, Jin R, Sudjianto A: **Analytical variance-based global sensitivity analysis in simulation-based design under uncertainty**. *J. Mech. Design* 2005, **127**:875–886.

42. Xu H, Rahman S: **A generalized dimension-reduction method for multidimensional integration in stochastic mechanics**. *Int. J. Numer. Meth. Eng.* 2004, **61**:1992–2019.

43. Xu H, Rahman S: **Decomposition methods for structural reliability analysis**. *Probabilist. Eng. Mech.* 2005, **20**:239–250.

44. Li G, Wang SW, Rabitz H: **Practical approaches to construct RS-HDMR component functions**. *J. Phys. Chem. A* 2002, **106**:8721–8733.

45. Wang SW, Georgopoulos PG, Li G, Rabitz H: **Random sampling-high dimensional model representation (RS-HDMR) with nonuniform distributed variables: Application to an integrated multimedia/multipathway exposure and dose model for trichloroethylene**. *J. Phys. Chem. A* 2003, **107**:4707–4716.

46. Sudret B: **Global sensitivity analysis using polynomial chaos expansions**. *Reliab. Eng. Syst. Safe.* 2008, **93**:964–979.

47. Crestaux T, Maître OL, Martinez JM: **Polynomial chaos expansion for sensitivity analysis**. *Reliab. Eng. Syst. Safe.* 2009, **94**:1161–1172.

48. Montgomery DC, Peck EA, Vining GG: *Introduction to Linear Regression Analysis*. New York: John Wiley, 3rd edition 2001.

49. Press WH, Teukolsky SA, Vetterling WT, Flannery BP: *Numerical Recipes: The Art of Scientific Computing*. New York: Cambridge University Press, 3rd edition 2007.

50. Aliş ÖF, Rabitz H: **Efficient implementation of high dimensional model representations**. *J. Math. Chem.* 2001, **29**:127–142.

51. Choi SK, Grandhi RV, Canfield RA, Pettit CL: **Polynomial chaos expansion with Latin hypercube sampling for estimating response variability**. *AIAA J.* 2004, **42**:1191–1198.

52. Castillo E, Sánchez-Maroño N, Alonso-Betanzos A, Castillo C: **Functional network topology learning and sensitivity analysis based on ANOVA decomposition**. *Neural. Comput.* 2007, **19**:231–257.

53. Li G, Hu J, Wang SW, Georgopoulos PG, Schoendorf J, Rabitz H: **Random sampling-high dimensional model represenattion (RS-HDMR) and orthogonality of its different order component functions**. *J. Phys. Chem. A* 2006, **110**:2474–2485.

Figure legends

Figure 1: A biochemical reaction model of the MAPK signaling cascade, adopted from Zhang *et al.* [16].

Tables

Table 1: Required system integrations, equations used, and sources of error.

Method	System Integrations	ROSA	SOSA	Equations Used	Error Sources
MC	$2L(J+1)$	264000	288000	(10)–(12)	• number of MC samples used
DA	$2J(J+1)+1$	925	1105	(14)–(16)	• local approximation • truncation of Taylor series • derivative approximation
PA	$J(J-1)S^2/2+JS+1$	3445	4141	(14), (15), (18)	• 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	(14), (15), (19)–(21)	• local approximation • truncation of FD-HDMR • Gauss-Hermite integration
OHA	L	6000	6000	(14), (15), (23)	• 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.

Table 2: ROSA-based sensitivity analysis results for the *duration* of ERK-PP activity.

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

Table 3: ROSA-based sensitivity analysis results for the *integrated response* of ERK-PP activity.

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

Table 4: ROSA-based sensitivity analysis results for the *strength* of ERK-PP activity.

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

Additional files

Additional file 1: Approximation methods and implementations. This file contains the mathematical details associated with the five approximation methods presented in the paper and discusses their numerical implementation.

Additional file 2: MAPK signaling cascade model. This file lists the biochemical reactions associated with the MAPK signaling cascade model and provides nominal values for the normalized reaction rate constants and initial molecular concentrations.

Additional file 3: SOSA-based sensitivity analysis results. This file summarizes 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 approximation methods discussed in the paper.

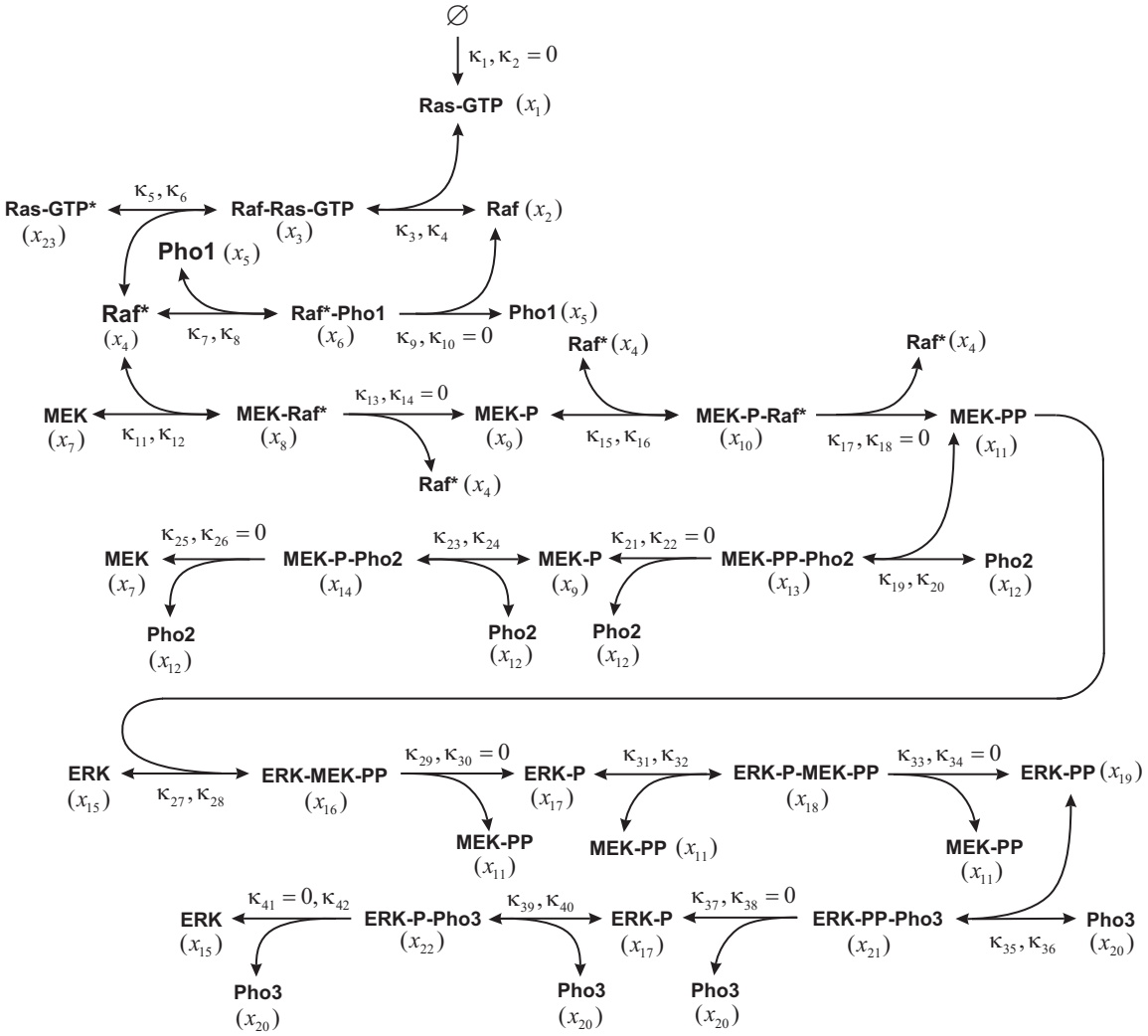
∅
κ1, κ2 = 0
Ras-GTP (x1)
Ras-GTP* (x23)
κ5, κ6
Raf-Ras-GTP (x3)
κ3, κ4
Raf (x2)
Pho1 (x5)
Raf* (x4)
κ7, κ8
Raf*-Pho1 (x6)
κ9, κ10 = 0
Pho1 (x5)
Raf* (x4)
Raf* (x4)
MEK (x7)
κ11, κ12
MEK-Raf* (x8)
κ13, κ14 = 0
MEK-P (x9)
κ15, κ16
MEK-P-Raf* (x10)
κ17, κ18 = 0
MEK-PP (x11)
Raf* (x4)
κ25, κ26 = 0
MEK (x7)
Pho2 (x12)
MEK-P-Pho2 (x14)
κ23, κ24
Pho2 (x12)
MEK-P (x9)
κ21, κ22 = 0
Pho2 (x12)
MEK-PP-Pho2 (x13)
κ19, κ20
Pho2 (x12)
ERK (x15)
κ27, κ28
ERK-MEK-PP (x16)
κ29, κ30 = 0
MEK-PP (x11)
ERK-P (x17)
κ31, κ32
MEK-PP (x11)
ERK-P-MEK-PP (x18)
κ33, κ34 = 0
MEK-PP (x11)
ERK-PP (x19)
ERK (x15)
κ41 = 0, κ42
Pho3 (x20)
ERK-P-Pho3 (x22)
κ39, κ40
Pho3 (x20)
ERK-P (x17)
κ37, κ38 = 0
Pho3 (x20)
ERK-PP-Pho3 (x21)
κ35, κ36
Pho3 (x20)

FIGURE 1

ADDITIONAL FILE 1

A comparison of approximation techniques for variance-based sensitivity analysis of biochemical reaction systems

APPROXIMATION METHODS AND IMPLEMENTATIONS

Hong-Xuan Zhang[1] **and John Goutsias**[*1]

[1] Whitaker Biomedical Engineering Institute, The Johns Hopkins University, Baltimore, MD 21218, USA

[*] Corresponding author

Email: HXZ: hxzhang@jhu.edu, JG: goutsias@jhu.edu

In this document, we present several methods for approximating the indices σ_j and η_j associated with the second-order variance-based sensitivity analysis technique discussed in the Main text. 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 σ_j and η_j. Finally, we present details regarding the numerical implementation of the resulting approximation techniques.

1 Response function representation schemes

For ease of presentation, we will often base our discussion on a response function $R(\boldsymbol{w}) = R(w_1, w_2, w_3)$ that depends only on three factors of interest $\boldsymbol{w} = \{w_1, w_2, w_3\}$. 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.

1.1 TSMR

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

$$\begin{aligned} R(w_1, w_2, w_3) = r_0 \; + \; & r_1(w_1) + r_2(w_2) + r_3(w_3) + r_{12}(w_1, w_2) \\ & + \; r_{13}(w_1, w_3) + r_{23}(w_2, w_3) + r_{123}(w_1, w_2, w_3), \end{aligned} \tag{S-1.1}$$

where

$$
\begin{aligned}
r_0 &:= R(\mathbf{0}) \\
r_1(w_1) &:= \sum_{m=1}^{\infty} \frac{1}{m!} \frac{\partial^m R(\mathbf{0})}{\partial w_1^m} \, w_1^m \\
r_2(w_2) &:= \sum_{m=1}^{\infty} \frac{1}{m!} \frac{\partial^m R(\mathbf{0})}{\partial w_2^m} \, w_2^m \\
r_3(w_3) &:= \sum_{m=1}^{\infty} \frac{1}{m!} \frac{\partial^m R(\mathbf{0})}{\partial w_3^m} \, w_3^m \\
r_{12}(w_1, w_2) &:= \sum_{m_1=1}^{\infty} \sum_{m_2=1}^{\infty} \frac{1}{m_1!\, m_2!} \frac{\partial^{m_1+m_2} R(\mathbf{0})}{\partial w_1^{m_1} \partial w_2^{m_2}} \, w_1^{m_1} w_2^{m_2} \\
r_{13}(w_1, w_3) &:= \sum_{m_1=1}^{\infty} \sum_{m_3=1}^{\infty} \frac{1}{m_1!\, m_3!} \frac{\partial^{m_1+m_3} R(\mathbf{0})}{\partial w_1^{m_1} \partial w_3^{m_3}} \, w_1^{m_1} w_3^{m_3} \\
r_{23}(w_2, w_3) &:= \sum_{m_2=1}^{\infty} \sum_{m_3=1}^{\infty} \frac{1}{m_2!\, m_3!} \frac{\partial^{m_2+m_3} R(\mathbf{0})}{\partial w_2^{m_2} \partial w_3^{m_3}} \, w_2^{m_2} w_3^{m_3} \\
r_{123}(w_1, w_2, w_3) &:= \sum_{m_1=1}^{\infty} \sum_{m_2=1}^{\infty} \sum_{m_3=1}^{\infty} \frac{1}{m_1!\, m_2!\, m_3!} \frac{\partial^{m_1+m_2+m_3} R(\mathbf{0})}{\partial w_1^{m_1} \partial w_2^{m_2} \partial w_3^{m_3}} \, w_1^{m_1} w_2^{m_2} w_3^{m_3}. \qquad \text{(S-1.2)}
\end{aligned}
$$

Clearly, the Taylor series expansion provides a representation of the system response R in terms of functions r, given by (S-1.2). 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, $r_1(w_1)$ summarizes the *singular* contribution of factor w_1 to the value of R, whereas, $r_{12}(w_1, w_2)$ summarizes the *joint* contribution of factors w_1 and w_2. Finally, $r_{123}(w_1, w_2, w_3)$ 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 (S-1.2) provides analytical formulas for the basis functions, calculating these functions at a point $\boldsymbol{w}$ 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 (S-1.2) is zero if one of its arguments equals zero. By using this property and (S-1.1), we have that

$$\begin{aligned}
R(w_1, w_2, 0) &= r_0 + r_1(w_1) + r_2(w_2) + r_{12}(w_1, w_2) \\
R(w_1, 0, w_3) &= r_0 + r_1(w_1) + r_3(w_3) + r_{13}(w_1, w_3) \\
R(0, w_2, w_3) &= r_0 + r_2(w_2) + r_3(w_3) + r_{23}(w_2, w_3) \\
R(w_1, 0, 0) &= r_0 + r_1(w_1) \\
R(0, w_2, 0) &= r_0 + r_2(w_2) \\
R(0, 0, w_3) &= r_0 + r_3(w_3) \\
R(0, 0, 0) &= r_0,
\end{aligned}$$

which results in

$$\begin{aligned}
r_0 &= R(0,0,0) \\
r_1(w_1) &= R(w_1,0,0) - R(0,0,0) \\
r_2(w_2) &= R(0,w_2,0) - R(0,0,0) \\
r_3(w_3) &= R(0,0,w_3) - R(0,0,0) \\
r_{12}(w_1,w_2) &= R(w_1,w_2,0) - R(w_1,0,0) - R(0,w_2,0) + R(0,0,0) \\
r_{13}(w_1,w_3) &= R(w_1,0,w_3) - R(w_1,0,0) - R(0,0,w_3) + R(0,0,0) \\
r_{23}(w_2,w_3) &= R(0,w_2,w_3) - R(0,w_2,0) - R(0,0,w_3) + R(0,0,0) \\
r_{123}(w_1,w_2,w_3) &= R(w_1,w_2,w_3) - R(w_1,w_2,0) - R(w_1,0,w_3) - R(0,w_2,w_3) \\
&\quad + R(w_1,0,0) + R(0,w_2,0) + R(0,0,w_3) \\
&\quad - R(0,0,0). \qquad \text{(S-1.3)}
\end{aligned}$$

These formulas provide a method for evaluating the basis functions r at some point $\boldsymbol{w}$. This can be done by calculating the system response at the corresponding w values suggested by the formulas. For example, evaluation of r_0 requires calculation of the system response at $w_1 = w_2 = w_3 = 0$, whereas, evaluation of $r_1(w_1)$ requires an additional calculation of the system response at $w_1, w_2 = w_3 = 0$. This can be done by solving the system of ordinary differential equations given by equations (1) and (2) in the Main text. We refer to the representation scheme given by (S-1.1) and (S-1.2) as *Taylor Series Model Representation* (TSMR).

1.2 FD-HDMR

We can extend the decomposition scheme given by (S-1.1) 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 $\boldsymbol{w} = \{w_1, w_2, \ldots, w_J\}$ by

$$R(\boldsymbol{w}) = r_0 + \sum_{j=1}^{J} \sum_{1 \le m_1 < \cdots < m_j \le J} \cdots \sum r_{m_1 m_2 \cdots m_j}(w_{m_1}, w_{m_2}, \ldots, w_{m_j}). \tag{S-1.4}$$

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 (S-1.4) is known in the literature as *High-Dimensional Model Representation* (HDMR) [1, 2].

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

$$r_{m_1 m_2 \cdots m_j}(w_{m_1}, w_{m_2}, \ldots, w_{m_j}) = \sum_{J \subseteq I} (-1)^{|I \setminus J|} R(\boldsymbol{w}_J), \tag{S-1.5}$$

which generalizes (S-1.3). 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 $\boldsymbol{w}_J$ is $\boldsymbol{w}$ with all variables, except the one indexed by J, set to zero.

Equations (S-1.4) and (S-1.5) express $R(\boldsymbol{w})$ 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 [1–3] or *Finite Difference* (FD) HDMR [4]. 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 (S-1.2).

1.3 ANOVA-HDMR

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

$$P(u_1, u_2, u_3) := R(g_1^{-1}(u_1), g_2^{-1}(u_2), g_3^{-1}(u_3)). \tag{S-1.6}$$

The HDMR representation of P is given by

$$P(u_1,u_2,u_3) = p_0 \; + \; p_1(u_1) + p_2(u_2) + p_3(u_3) + p_{12}(u_1,u_2) \\ + \; p_{13}(u_1,u_3) + p_{23}(u_2,u_3) + p_{123}(u_1,u_2,u_3). \tag{S-1.7}$$

If we consider basis functions p that integrate to zero over a *single* variable, then we can readily verify from (S-1.7) that

$$\begin{aligned}
p_0 &= \int_0^1\int_0^1\int_0^1 P(u_1,u_2,u_3)du_1du_2du_3 \\
p_1(u_1) &= \int_0^1\int_0^1 P(u_1,u_2,u_3)du_2du_3 - p_0 \\
p_2(u_2) &= \int_0^1\int_0^1 P(u_1,u_2,u_3)du_1du_3 - p_0 \\
p_3(u_3) &= \int_0^1\int_0^1 P(u_1,u_2,u_3)du_1du_2 - p_0 \\
p_{12}(u_1,u_2) &= \int_0^1 P(u_1,u_2,u_3)du_3 - p_1(u_1) - p_2(u_2) - p_0 \\
p_{13}(u_1,u_3) &= \int_0^1 P(u_1,u_2,u_3)du_2 - p_1(u_1) - p_3(u_3) - p_0 \\
p_{23}(u_2,u_3) &= \int_0^1 P(u_1,u_2,u_3)du_1 - p_2(u_2) - p_3(u_3) - p_0 \\
p_{123}(u_1,u_2,u_3) &= P(u_1,u_2,u_3) - p_{12}(u_1,u_2) - p_{13}(u_1,u_3) - p_{23}(u_2,u_3) \\
&\quad -p_1(u_1) - p_2(u_2) - p_3(u_3) - p_0.
\end{aligned} \tag{S-1.8}$$

Therefore, we can uniquely determine the basis functions p from P. By setting $u_j = g_j(w_j)$ in (S-1.7) and (S-1.8), and by employing (S-1.6), we obtain

$$R(w_1,w_2,w_3) = \rho_0 \; + \; \rho_1(w_1) + \rho_2(w_2) + \rho_3(w_3) + \rho_{12}(w_1,w_2) \\ + \; \rho_{13}(w_1,w_3) + \rho_{23}(w_2,w_3) + \rho_{123}(w_1,w_2,w_3), \tag{S-1.9}$$

where

$$\rho_0 := p_0 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R(w_1,w_2,w_3)g_1'(w_1)g_2'(w_2)g_3'(w_3)dw_1dw_2dw_3$$

$$\rho_1(w_1) := p_1(g_1(w_1)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R(w_1,w_2,w_3)g_2'(w_2)g_3'(w_3)dw_2dw_3 - \rho_0$$

$$\rho_2(w_2) := p_2(g_2(w_2)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R(w_1,w_2,w_3)g_1'(w_1)g_3'(w_3)dw_1dw_3 - \rho_0$$

$$\rho_3(w_3) := p_3(g_3(w_3)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} R(w_1,w_2,w_3)g_1'(w_1)g_2'(w_2)dw_1dw_2 - \rho_0$$

$$\rho_{12}(w_1,w_2) := p_{12}(g_1(w_1),g_2(w_2)) = \int_{-\infty}^{\infty} R(w_1,w_2,w_3)g_3'(w_3)dw_3 - \rho_1(w_1) - \rho_2(w_2) - \rho_0$$

$$\rho_{13}(w_1,w_3) := p_{13}(g_1(w_1),g_3(w_3)) = \int_{-\infty}^{\infty} R(w_1,w_2,w_3)g_2'(w_2)dw_2 - \rho_1(w_1) - \rho_3(w_3) - \rho_0$$

$$\rho_{23}(w_2,w_3) := p_{23}(g_2(w_2),g_3(w_3)) = \int_{-\infty}^{\infty} R(w_1,w_2,w_3)g_1'(w_1)dw_1 - \rho_2(w_2) - \rho_3(w_3) - \rho_0$$

$$\begin{aligned}\rho_{123}(w_1,w_2,w_3) &:= p_{123}(g_1(w_1),g_2(w_2),g_3(w_3))\\ &= R(w_1,w_2,w_3) - \rho_{12}(w_1,w_2) - \rho_{13}(w_1,w_3) - \rho_{23}(w_2,w_3)\\ &\quad -\rho_1(w_1) - \rho_2(w_2) - \rho_3(w_3) - \rho_0,\end{aligned} \tag{S-1.10}$$

with g' being the first-order derivative of g. For reasons to be explained in Section 2.3, the representation of a response function R by (S-1.9) and (S-1.10) is referred to in the literature as *Analysis-of-Variance* (ANOVA) HDMR [1–5]. Note that the basis functions ρ satisfy the following orthogonality conditions:

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \rho_{j_1,\ldots,j_k}(w_{j_1},\ldots,w_{j_k})g_1'(w_1)\cdots g_J'(w_J)dw_1\cdots dw_J = 0,$$

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \rho_{j_1,\ldots,j_k}(w_{j_1},\ldots,w_{j_k})\rho_{j_1',\ldots,j_{k'}'}(w_{j_1'},\ldots,w_{j_{k'}'})g_1'(w_1)\cdots g_J'(w_J)dw_1\cdots dw_J = 0,$$
$$(j_1,\ldots,j_k) \neq (j_1',\ldots,j_{k'}'), \tag{S-1.11}$$

provided that the derivatives $g_j'(w_j)$ integrate to one.

2 Approximation of response variances

In this section, we assume that the biochemical factors of interest are statistically independent random variables $W_1, W_2, \ldots, W_J$ that follow zero-mean Gaussian distributions with standard deviations

$\lambda_1, \lambda_2, \ldots, \lambda_J$, respectively. In this case, the response $R(\boldsymbol{W})$ of the biochemical reaction system, where $\boldsymbol{W} = \{W_1, W_2, \ldots, W_J\}$, will be a random variable as well. We are interested in evaluating the following response variances:

$$\begin{aligned} V_j &:= \text{Var}[\text{E}[R(\boldsymbol{W}) \mid W_j]] \\ V_{jj'} &:= \text{Var}[\text{E}[R(\boldsymbol{W}) \mid W_j, W_{j'}]] - V_j - V_{j'}. \end{aligned} \tag{S-1.12}$$

We can then calculate the (second-order) SESI's and JESI's by means of

$$\sigma_j = \frac{V_j}{V} \qquad \text{and} \qquad \eta_j = \frac{U_j}{V}\,, \tag{S-1.13}$$

where

$$U_j := \sum_{j'=1, j' \neq j}^{J} V_{jj'} \qquad \text{and} \qquad V := \sum_{j=1}^{J} V_j + \sum_{j=1}^{J-1} \sum_{j'=j+1}^{J} V_{jj'}. \tag{S-1.14}$$

In most applications of interest however it is very difficult to evaluate the previous variances due to the complexity of the response function R. We can address this problem by replacing the response function with a simpler function $\widehat{R}(w_1, w_2, \ldots, w_J)$ that will allow us to approximate the response variances given by (S-1.12). In the following, we discuss various approximations obtained by employing the previously discussed representation schemes.

2.1 TSMR

As we mentioned in Section 1.1, the two main problems associated with the basis functions of TSMR, given by (S-1.2), is the need to calculate high-order partial derivatives of the response function and 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 order greater than two are negligible. In this case, we can approximate the response function $R(w_1, w_2, w_3)$ by

$$\widehat{R}(w_1, w_2, w_3) = \widehat{r}_0 + \widehat{r}_1(w_1) + \widehat{r}_2(w_2) + \widehat{r}_3(w_3) + \widehat{r}_{12}(w_1, w_2) + \widehat{r}_{13}(w_1, w_3) + \widehat{r}_{23}(w_2, w_3),$$

where

$$\begin{aligned} \widehat{r}_0 &:= R(\mathbf{0}) \\ \widehat{r}_1(w_1) &:= \frac{\partial R(\mathbf{0})}{\partial w_1}\, w_1 + \frac{1}{2} \frac{\partial^2 R(\mathbf{0})}{\partial w_1^2}\, w_1^2 \\ \widehat{r}_2(w_2) &:= \frac{\partial R(\mathbf{0})}{\partial w_2}\, w_2 + \frac{1}{2} \frac{\partial^2 R(\mathbf{0})}{\partial w_2^2}\, w_2^2 \end{aligned}$$

$$\widehat{r}_3(w_3) := \frac{\partial R(\mathbf{0})}{\partial w_3}\, w_3 + \frac{1}{2}\frac{\partial^2 R(\mathbf{0})}{\partial w_3^2}\, w_3^2$$

$$\widehat{r}_{12}(w_1, w_2) := \frac{\partial^2 R(\mathbf{0})}{\partial w_1 \partial w_2}\, w_1 w_2$$

$$\widehat{r}_{13}(w_1, w_3) := \frac{\partial^2 R(\mathbf{0})}{\partial w_1 \partial w_3}\, w_1 w_3$$

$$\widehat{r}_{23}(w_2, w_3) := \frac{\partial^2 R(\mathbf{0})}{\partial w_2 \partial w_3}\, w_2 w_3\,,$$

since $r_{123}(w_1, w_2, w_3) = 0$ in this case. By employing the statistical independence of W_1, W_2, and W_3, we can show that the variances associated with the approximate response function $\widehat{R}(w_1, w_2, w_3)$ satisfy:

$$\begin{aligned}
\widehat{V} &= d_1^2\lambda_1^2 + d_2^2\lambda_2^2 + d_3^2\lambda_3^2 + \frac{1}{2}\, d_{11}^2\lambda_1^4 + \frac{1}{2}\, d_{22}^2\lambda_2^4 + \frac{1}{2}\, d_{33}^2\lambda_3^4 + d_{12}^2\lambda_1^2\lambda_2^2 + d_{13}^2\lambda_1^2\lambda_3^2 + d_{23}^2\lambda_2^2\lambda_3^2 \\
\widehat{V}_1 &= d_1^2\lambda_1^2 + \frac{1}{2}\, d_{11}^2\lambda_1^4 \\
\widehat{V}_2 &= d_2^2\lambda_2^2 + \frac{1}{2}\, d_{22}^2\lambda_2^4 \\
\widehat{V}_3 &= d_3^2\lambda_3^2 + \frac{1}{2}\, d_{33}^2\lambda_3^4 \\
\widehat{V}_{12} &= d_{12}^2\lambda_1^2\lambda_2^2 \\
\widehat{V}_{13} &= d_{13}^2\lambda_1^2\lambda_3^2 \\
\widehat{V}_{23} &= d_{23}^2\lambda_2^2\lambda_3^2\,,
\end{aligned} \tag{S-1.15}$$

where d_j is the first-order partial derivative of R with respect to w_j at $\mathbf{0}$ and $d_{jj'}$ is the second-order partial derivative of R with respect to w_j and $w_{j'}$ at $\mathbf{0}$. To show (S-1.15), we have used the fact that W_j follows a Gaussian distribution with zero mean and standard deviation λ_j, which implies $\mathrm{E}[W_j^3] = 0$, $\mathrm{E}[W_j^4] = 3\lambda_j^4$. As a consequence of (S-1.13), (S-1.14), and (S-1.15), we obtain the following approximations to the SESI's and JESI's (expressed for the general case of J biochemical factors):

$$\boxed{\begin{aligned}
\widehat{\sigma}_j &= \frac{\widehat{V}_j}{\widehat{V}}, \qquad \widehat{\eta}_j = \frac{\widehat{U}_j}{\widehat{V}} \\
\widehat{V}_j &= \lambda_j^2 d_j^2 + \frac{1}{2}\,\lambda_j^4 d_{jj}^2 \\
\widehat{V}_{jj'} &= \lambda_j^2\lambda_{j'}^2 d_{jj'}^2 \\
\widehat{U}_j &= \sum_{j'=1, j'\neq j}^{J} \widehat{V}_{jj'} \\
\widehat{V} &= \sum_{j=1}^{J} \widehat{V}_j + \sum_{j=1}^{J-1}\sum_{j'=j+1}^{J} \widehat{V}_{jj'}
\end{aligned}} \tag{S-1.16}$$

We respectively refer to $\widehat{\sigma}_j$ and $\widehat{\eta}_j$, given by (S-1.16), as the SESI's and JESI's obtained by *Derivative Approximation* (DA).

2.2 FD-HDMR

2.2.1 Polynomial approximation

We may obtain a better approximation of the sensitivity indices σ_j and η_j by assuming that the response function is sufficiently smooth in a neighborhood around $\mathbf{0}$ so that partial derivatives of order 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 $R(w_1, w_2, w_3)$ by

$$\widehat{R}(w_1, w_2, w_3) = \widehat{r}_0 + \widehat{r}_1(w_1) + \widehat{r}_2(w_2) + \widehat{r}_3(w_3) + \widehat{r}_{12}(w_1, w_2) + \widehat{r}_{13}(w_1, w_3) + \widehat{r}_{23}(w_2, w_3), \quad \text{(S-1.17)}$$

where

$$\begin{aligned}
\widehat{r}_0 &:= R(\mathbf{0}) \\
\widehat{r}_1(w_1) &:= \frac{\partial R(\mathbf{0})}{\partial w_1}\, w_1 + \frac{1}{2}\frac{\partial^2 R(\mathbf{0})}{\partial w_1^2}\, w_1^2 \\
\widehat{r}_2(w_2) &:= \frac{\partial R(\mathbf{0})}{\partial w_2}\, w_2 + \frac{1}{2}\frac{\partial^2 R(\mathbf{0})}{\partial w_2^2}\, w_2^2 \\
\widehat{r}_3(w_3) &:= \frac{\partial R(\mathbf{0})}{\partial w_3}\, w_3 + \frac{1}{2}\frac{\partial^2 R(\mathbf{0})}{\partial w_3^2}\, w_3^2 \\
\widehat{r}_{12}(w_1, w_2) &:= \frac{\partial^2 R(\mathbf{0})}{\partial w_1 \partial w_2}\, w_1 w_2 + \frac{1}{2}\frac{\partial^3 R(\mathbf{0})}{\partial w_1^2 \partial w_2}\, w_1^2 w_2 + \frac{1}{2}\frac{\partial^3 R(\mathbf{0})}{\partial w_1 \partial w_2^2}\, w_1 w_2^2 + \frac{1}{4}\frac{\partial^4 R(\mathbf{0})}{\partial w_1^2 \partial w_2^2}\, w_1^2 w_2^2 \\
\widehat{r}_{13}(w_1, w_3) &:= \frac{\partial^2 R(\mathbf{0})}{\partial w_1 \partial w_3}\, w_1 w_3 + \frac{1}{2}\frac{\partial^3 R(\mathbf{0})}{\partial w_1^2 \partial w_3}\, w_1^2 w_3 + \frac{1}{2}\frac{\partial^3 R(\mathbf{0})}{\partial w_1 \partial w_3^2}\, w_1 w_3^2 + \frac{1}{4}\frac{\partial^4 R(\mathbf{0})}{\partial w_1^2 \partial w_3^2}\, w_1^2 w_3^2 \\
\widehat{r}_{23}(w_2, w_3) &:= \frac{\partial^2 R(\mathbf{0})}{\partial w_2 \partial w_3}\, w_2 w_3 + \frac{1}{2}\frac{\partial^3 R(\mathbf{0})}{\partial w_2^2 \partial w_3}\, w_2^2 w_3 + \frac{1}{2}\frac{\partial^3 R(\mathbf{0})}{\partial w_2 \partial w_3^2}\, w_2 w_3^2 + \frac{1}{4}\frac{\partial^4 R(\mathbf{0})}{\partial w_2^2 \partial w_3^2}\, w_2^2 w_3^2. \quad \text{(S-1.18)}
\end{aligned}$$

Due to difficulties in numerically evaluating high-order derivatives with sufficient accuracy, we may not be able to use (S-1.18) to derive sufficiently good DA approximations of the sensitivity indices. However, this equation motivates us to set

$$\begin{aligned}
\widehat{r}_j(w_j) &= \alpha_{j,1} w_j + \alpha_{j,2}\, w_j^2 \\
\widehat{r}_{jj'}(w_j, w_{j'}) &= \alpha_{jj',1}\, w_j w_{j'} + \alpha_{jj',2}\, w_j^2 w_{j'} + \alpha_{jj',3}\, w_j w_{j'}^2 + \alpha_{jj',4}\, w_j^2 w_{j'}^2, \quad \text{(S-1.19)}
\end{aligned}$$

where the α's are parameters whose values must be appropriately determined so that $\widehat{R}$, given by (S-1.17) and (S-1.19), sufficiently approximates the response function R. We will be discussing a practical method to address this problem in Section 3 of this supplement.

Clearly, the previous approach is based on approximating the first- and second-order basis functions associated with the FD-HDMR given by (S-1.4) with the polynomials given by (S-1.19). If $\widehat{R}(\boldsymbol{w})$ is sufficiently close to $R(\boldsymbol{w})$ in a neighborhood around $\mathbf{0}$, then the parameters α will coincide with the partial derivatives of R associated with (S-1.18). Note that the approximating basis functions $\widehat{r}$ given by (S-1.19) satisfy the necessary condition of becoming zero if one of their arguments equals zero.

As a consequence of (S-1.13) and (S-1.14), by employing the statistical independence of the W_j's, and by using the fact that W_j follows a Gaussian distribution with zero mean and standard deviation λ_j, in which case $\mathrm{E}[W_j^3] = 0$ and $\mathrm{E}[W_j^4] = 3\lambda_j^4$, we obtain the following approximations to the SESI's and JESI's:

$$
\begin{aligned}
\widehat{\sigma}_j &= \frac{\widehat{V}_j}{\widehat{V}}, \qquad \widehat{\eta}_j = \frac{\widehat{U}_j}{\widehat{V}} \\
\widehat{V}_j &= \lambda_j^2\alpha_{j,1}^2 + 2\lambda_j^4\alpha_{j,2}^2 + 2\lambda_j^2\alpha_{j,1}\left(\sum_{m=1}^{j-1}\lambda_m^2\alpha_{mj,2} + \sum_{m=j+1}^{J}\lambda_m^2\alpha_{jm,3}\right) \\
&\quad + \lambda_j^2\left(\sum_{m=1}^{j-1}\lambda_m^2\alpha_{mj,2} + \sum_{m=j+1}^{J}\lambda_m^2\alpha_{jm,3}\right)^2 \\
&\quad + 4\lambda_j^4\alpha_{j,2}\left(\sum_{m=1}^{j-1}\lambda_m^2\alpha_{mj,4} + \sum_{m=j+1}^{J}\lambda_m^2\alpha_{jm,4}\right) \\
&\quad + 2\lambda_j^4\left(\sum_{m=1}^{j-1}\lambda_m^2\alpha_{mj,4} + \sum_{m=j+1}^{J}\lambda_m^2\alpha_{jm,4}\right)^2 \\
\widehat{V}_{jj'} &= \lambda_j^2\lambda_{j'}^2\alpha_{jj',1}^2 + 2\lambda_j^4\lambda_{j'}^2\alpha_{jj',2}^2 + 2\lambda_j^2\lambda_{j'}^4\alpha_{jj',3}^2 + 4\lambda_j^4\lambda_{j'}^4\alpha_{jj',4}^2 \\
\widehat{U}_j &= \sum_{j'=1,j'\neq j}^{J}\widehat{V}_{jj'} \\
\widehat{V} &= \sum_{j=1}^{J}\widehat{V}_j + \sum_{j=1}^{J-1}\sum_{j'=j+1}^{J}\widehat{V}_{jj'}
\end{aligned}
\tag{S-1.20}
$$

Note that (S-1.20) is a special case of equations 35 and 36 in [6]. We respectively refer to $\widehat{\sigma}_j$ and $\widehat{\eta}_j$, given by (S-1.20), as the SESI's and JESI's obtained by *Polynomial Approximation* (PA) of the FD-HDMR.

2.2.2 Gauss-Hermite integration

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 $R(w_1, w_2, w_3)$ by

$$\widehat{R}(w_1, w_2, w_3) = r_0 + r_1(w_1) + r_2(w_2) + r_3(w_3) + r_{12}(w_1, w_2) + r_{13}(w_1, w_3) + r_{23}(w_2, w_3), \quad \text{(S-1.21)}$$

where

$$\begin{aligned}
r_0 &= R(\mathbf{0}) \\
r_1(w_1) &= \sum_{m=1}^{\infty} \frac{1}{m!} \frac{\partial^m R(\mathbf{0})}{\partial w_1^m} w_1^m \\
r_2(w_2) &= \sum_{m=1}^{\infty} \frac{1}{m!} \frac{\partial^m R(\mathbf{0})}{\partial w_2^m} w_2^m \\
r_3(w_3) &= \sum_{m=1}^{\infty} \frac{1}{m!} \frac{\partial^m R(\mathbf{0})}{\partial w_3^m} w_3^m \\
r_{12}(w_1, w_2) &= \sum_{m_1=1}^{\infty} \sum_{m_2=1}^{\infty} \frac{1}{m_1! m_2!} \frac{\partial^{m_1+m_2} R(\mathbf{0})}{\partial w_1^{m_1} \partial w_2^{m_2}} w_1^{m_1} w_2^{m_2} \\
r_{13}(w_1, w_3) &= \sum_{m_1=1}^{\infty} \sum_{m_3=1}^{\infty} \frac{1}{m_1! m_3!} \frac{\partial^{m_1+m_3} R(\mathbf{0})}{\partial w_1^{m_1} \partial w_3^{m_3}} w_1^{m_1} w_3^{m_3} \\
r_{23}(w_2, w_3) &= \sum_{m_2=1}^{\infty} \sum_{m_3=1}^{\infty} \frac{1}{m_2! m_3!} \frac{\partial^{m_2+m_3} R(\mathbf{0})}{\partial w_2^{m_2} \partial w_3^{m_3}} w_2^{m_2} w_3^{m_3}, \qquad \text{(S-1.22)}
\end{aligned}$$

since $r_{123}(w_1, w_2, w_3) = 0$ in this case. We expect that this approximation will be more accurate than the one considered in (S-1.17) and (S-1.18), since the first- and second-order basis functions are exactly the same as the corresponding basis functions given by (S-1.2). Note that we can obtain the approximation given by (S-1.21) by simply truncating the third- and higher-order terms in the FD-HDMR of the response function R, given by (S-1.4), without making any reference to the derivatives of R.

Since the basis functions r given by(S-1.22) become zero if one of their arguments is zero, we can relate them to the system response R by means of (S-1.3). As a consequence of (S-1.3) and (S-1.21), we obtain

the following decomposition for $\widehat{R}$ (expressed for the general case of J biochemical factors):

$$\widehat{R}(\boldsymbol{w}) = \psi_0 - (J-2)\sum_{j=1}^{J}\psi_j(w_j) + \sum_{j=1}^{J-1}\sum_{j'=j+1}^{J}\psi_{jj'}(w_j, w_{j'}), \tag{S-1.23}$$

where

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

By taking conditional and unconditional expectations on both sides of (S-1.23), and by using the statistical independence of the biochemical factors, we obtain

$$\begin{aligned}
e_0 &:= \mathrm{E}[\widehat{R}(\boldsymbol{W})] \\
&= \psi_0 - (J-2)\sum_{m=1}^{J}\mathrm{E}[\psi_m(W_m)] + \sum_{m=1}^{J-1}\sum_{m'=m+1}^{J}\mathrm{E}[\psi_{mm'}(W_m,W_{m'})] \\
e_j(w_j) &:= \mathrm{E}[\widehat{R}(\boldsymbol{W}) \mid W_j = w_j] \\
&= \psi_0 - (J-2)\sum_{m=1}^{J}\mathrm{E}[\psi_m(W_m) \mid W_j = w_j] \\
&\qquad + \sum_{m=1}^{J-1}\sum_{m'=m+1}^{J}\mathrm{E}[\psi_{mm'}(W_m,W_{m'}) \mid W_j = w_j] \\
e_{jj'}(w_j,w_{j'}) &:= \mathrm{E}[R(\boldsymbol{W}) \mid W_j = w_j, W_{j'} = w_{j'}] \\
&= \psi_0 - (J-2)\sum_{m=1}^{J}\mathrm{E}[\psi_m(W_m) \mid W_j = w_j, W_{j'} = w_{j'}] \\
&\qquad + \sum_{m=1}^{J-1}\sum_{m'=m+1}^{J}\mathrm{E}[\psi_{mm'}(W_m,W_{m'}) \mid W_j = w_j, W_{j'} = w_{j'}],
\end{aligned} \tag{S-1.24}$$

where

$$\mathrm{E}[\psi_m(W_m) \mid W_j = w_j] \;=\; \begin{cases} \psi_j(w_j), & \text{if } \; m = j \\ \mathrm{E}[\psi_m(W_m)], & \text{otherwise} \end{cases}$$

$$\mathrm{E}[\psi_m(W_m) \mid W_j = w_j, W_{j'} = w_{j'}] \;=\; \begin{cases} \psi_j(w_j), & \text{if } \; m = j \\ \psi_{j'}(w_{j'}), & \text{if } \; m = j' \,, \quad \text{for } \; j < j' \\ \mathrm{E}[\psi_m(W_m)], & \text{otherwise} \end{cases}$$

$$\mathrm{E}[\psi_{mm'}(W_m, W_{m'}) \mid W_j = w_j] \;=\; \begin{cases} \mathrm{E}[\psi_{mj}(W_m, w_j)], & \text{if } \; m < j, m' = j \\ \mathrm{E}[\psi_{jm'}(w_j, W_{m'})], & \text{if } \; m = j, m' > j \\ \mathrm{E}[\psi_{mm'}(W_m, W_{m'})], & \text{otherwise} \end{cases}$$

$$\mathrm{E}[\psi_{mm'}(W_m, W_{m'}) \mid W_j = w_j, W_{j'} = w_{j'}] \;=\; \begin{cases} \psi_{jj'}(w_j, w_{j'}), & \text{if } \; m = j, m' = j' \\ \mathrm{E}[\psi_{mj}(W_m, w_j)], & \text{if } \; m < j, m' = j \\ \mathrm{E}[\psi_{mj'}(W_m, w_{j'})], & \text{if } \; m \neq j, m' = j' \\ \mathrm{E}[\psi_{jm'}(w_j, W_{m'})], & \text{if } \; m = j, m' \neq j' \,, \\ \mathrm{E}[\psi_{j'm'}(w_{j'}, W_{m'})], & \text{if } \; m = j', m' > j' \\ \mathrm{E}[\psi_{mm'}(W_m, W_{m'})], & \text{otherwise} \end{cases} \quad \text{for } \; j < j'.$$

(S-1.25)

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

$$\begin{aligned} \mathrm{Var}[\mathrm{E}[\widehat{R}(\boldsymbol{W}) \mid W_j]] \;&=\; \mathrm{Var}[e_j(W_j)] = \mathrm{E}[e_j^2(W_j)] - e_0^2 \\ \mathrm{Var}[\mathrm{E}[\widehat{R}(\boldsymbol{W}) \mid W_j, W_{j'}]] \;&=\; \mathrm{Var}[e_{jj'}(W_j, W_{j'})] = \mathrm{E}[e_{jj'}^2(W_j, W_{j'})] - e_0^2, \end{aligned} \tag{S-1.26}$$

since

$$\begin{aligned} \mathrm{E}[e_j(W_j)] \;&=\; \mathrm{E}[\mathrm{E}[\widehat{R}(\boldsymbol{W}) \mid W_j]] = \mathrm{E}[\widehat{R}(\boldsymbol{W})] \\ \mathrm{E}[e_{jj'}(W_j, W_{j'})] \;&=\; \mathrm{E}[\mathrm{E}[\widehat{R}(\boldsymbol{W}) \mid W_j, W_{j'}]] = \mathrm{E}[\widehat{R}(\boldsymbol{W})], \end{aligned}$$

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

As a consequence of (S-1.13), (S-1.14), and (S-1.26), we now obtain the following approximations to the SESI's and JESI's:

$$\boxed{\begin{gathered}\widehat{\sigma}_j = \frac{\widehat{V}_j}{\widehat{V}}, \qquad \widehat{\eta}_j = \frac{\widehat{U}_j}{\widehat{V}} \\ \widehat{V}_j = \mathrm{E}[e_j^2(W_j)] - e_0^2 \\ \widehat{V}_{jj'} = \mathrm{E}[e_{jj'}^2(W_j, W_{j'})] - \widehat{V}_j - \widehat{V}_{j'} - e_0^2 \\ \widehat{U}_j = \sum_{j'=1, j' \neq j}^{J} \widehat{V}_{jj'} \\ \widehat{V} = \sum_{j=1}^{J} \widehat{V}_j + \sum_{j=1}^{J-1} \sum_{j'=j+1}^{J} \widehat{V}_{jj'}\end{gathered}} \tag{S-1.27}$$

with e_0, e_j, and $e_{jj'}$ given by (S-1.24). 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 Section 3. We respectively refer to $\widehat{\sigma}_j$ and $\widehat{\eta}_j$, given by Equation S-1.27, as the SESI's and JESI's obtained by *Gauss-Hermite Integration* (GHI) of the FD-HDMR.

2.3 ANOVA-HDMR

Equation (S-1.7) and the fact that the basis functions p integrate to zero over a single variable imply

$$\begin{aligned}\int_0^1 \int_0^1 \int_0^1 P^2(u_1, u_2, u_3) du_1 du_2 du_3 = p_0^2 \; &+ \int_0^1 p_1^2(u_1) du_1 + \int_0^1 p_2^2(u_2) du_2 + \int_0^1 p_3^2(u_3) du_3 \\ &+ \int_0^1 \int_0^1 p_{12}^2(u_1, u_2) du_1 du_2 \\ &+ \int_0^1 \int_0^1 p_{13}^2(u_1, u_3) du_1 du_3 \\ &+ \int_0^1 \int_0^1 p_{23}^2(u_2, u_3) du_2 du_3 \\ &+ \int_0^1 \int_0^1 \int_0^1 p_{123}^2(u_1, u_2, u_3) du_1 du_2 du_3 \, .\end{aligned} \tag{S-1.28}$$

If we assume that the biochemical factors of interest are statistically independent random variables W_1, W_2, and W_3, with cumulative distribution functions $g_1(w_1)$, $g_2(w_2)$, and $g_3(w_3)$, respectively, then (S-1.28), together with (S-1.6), (S-1.8), and (S-1.10), implies that

$$V = V_1 + V_2 + V_3 + V_{12} + V_{13} + V_{23} + V_{123}, \tag{S-1.29}$$

where

$$
\begin{aligned}
V &:= \text{Var}[R(W_1, W_2, W_3)] \\
V_1 &:= \text{Var}[\text{E}[R(W_1, W_2, W_3) \mid W_1]] = \int_{-\infty}^{\infty} \rho_1^2(w_1) g_1'(w_1) dw_1 \\
V_2 &:= \text{Var}[\text{E}[R(W_1, W_2, W_3) \mid W_2]] = \int_{-\infty}^{\infty} \rho_2^2(w_2) g_2'(w_2) dw_2 \\
V_3 &:= \text{Var}[\text{E}[R(W_1, W_2, W_3) \mid W_3]] = \int_{-\infty}^{\infty} \rho_3^2(w_3) g_3'(w_3) dw_3 \\
V_{12} &:= \text{Var}[\text{E}[R(W_1, W_2, W_3) \mid W_1, W_2]] - V_1 - V_2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \rho_{12}^2(w_1, w_2) g_1'(w_1) g_2'(w_2) dw_1 dw_2 \geq 0 \\
V_{13} &:= \text{Var}[\text{E}[R(W_1, W_2, W_3) \mid W_1, W_3]] - V_1 - V_3 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \rho_{13}^2(w_1, w_3) g_1'(w_1) g_3'(w_3) dw_1 dw_3 \geq 0 \\
V_{23} &:= \text{Var}[\text{E}[R(W_1, W_2, W_3) \mid W_2, W_3]] - V_2 - V_3 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \rho_{23}^2(w_2, w_3) g_2'(w_2) g_3'(w_3) dw_2 dw_3 \geq 0 \\
V_{123} &:= V - V_{12} - V_{13} - V_{23} - V_1 - V_2 - V_3 \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \rho_{123}^2(w_1, w_2, w_3) g_1'(w_1) g_2'(w_2) g_3'(w_3) dw_1 dw_2 dw_3 \geq 0, \qquad \text{(S-1.30)}
\end{aligned}
$$

since $g_1'(w_1)$, $g_2'(w_2)$, and $g_3'(w_3)$ are the probability density functions of W_1, W_2, and W_3, respectively. The variance decomposition scheme given by Equations 7 and 8 in the Main text is a general version of the decomposition given by (S-1.29) and (S-1.30) for the case of J biochemical factors. This decomposition is closely related to *analysis of variance* (ANOVA) techniques in statistics [5, 7, 8]. For this reason, the representation of the response function R by (S-1.9) and (S-1.10) is referred to in the literature as ANOVA-HDMR.

Note that (S-1.29) can be shown in a trivial manner by adding all V's in (S-1.30). However, by using (S-1.6), (S-1.8), (S-1.10), and (S-1.28), we can show that, when W_1, W_2, and W_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 paper. Moreover, we can show that these quantities can be directly evaluated from the basis functions of the ANOVA-HDMR of the response function $R(\boldsymbol{w})$ by means of (S-1.30). As a consequence, we can use ANOVA-HDMR to develop an efficient approximation technique for the sensitivity indices σ_j and η_j. We can do this by sufficiently approximating the response $R(w_1, w_2, w_3)$ by a function

$$
\widehat{R}(w_1, w_2, w_3) = \widehat{\rho}_0 + \widehat{\rho}_1(w_1) + \widehat{\rho}_2(w_2) + \widehat{\rho}_3(w_3) + \widehat{\rho}_{12}(w_1, w_2) + \widehat{\rho}_{13}(w_1, w_3) + \widehat{\rho}_{23}(w_2, w_3), \qquad \text{(S-1.31)}
$$

where the approximating basis functions $\widehat{\rho}$ must be appropriately chosen so that they satisfy the necessary

orthogonality conditions, given by (S-1.11), and allow efficient evaluation of the integrals in (S-1.30).

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:

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

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

$$\begin{aligned}
&\int_{-\infty}^{\infty} H_1(x) \, \frac{1}{\sqrt{2\pi}} \, e^{-x^2/2} dx = \int_{-\infty}^{\infty} H_2(x) \, \frac{1}{\sqrt{2\pi}} \, e^{-x^2/2} dx = 0 \\
&\int_{-\infty}^{\infty} H_1^2(x) \, \frac{1}{\sqrt{2\pi}} \, e^{-x^2/2} dx = \int_{-\infty}^{\infty} H_2^2(x) \, \frac{1}{\sqrt{2\pi}} \, e^{-x^2/2} dx = 1 \\
&\int_{-\infty}^{\infty} H_1(x) H_2(x) \, \frac{1}{\sqrt{2\pi}} \, e^{-x^2/2} dx = 0.
\end{aligned} \tag{S-1.32}$$

In this case, we set

$$\begin{aligned}
\widehat{\rho}_j(w_j) \;&=\; \alpha_{j,1} \frac{w_j}{\lambda_j} + \frac{\alpha_{j,2}}{\sqrt{2}} \left(\frac{w_j^2}{\lambda_j^2} - 1 \right) \\
\widehat{\rho}_{jj'}(w_j, w_{j'}) \;&=\; \alpha_{jj',1} \frac{w_j w_{j'}}{\lambda_j \lambda_{j'}} + \frac{\alpha_{jj',2}}{\sqrt{2}} \left(\frac{w_j^2}{\lambda_j^2} - 1 \right) \frac{w_{j'}}{\lambda_{j'}} + \frac{\alpha_{jj',3}}{\sqrt{2}} \frac{w_j}{\lambda_j} \left(\frac{w_{j'}^2}{\lambda_{j'}^2} - 1 \right) \\
&\qquad + \frac{\alpha_{jj',4}}{2} \left(\frac{w_j^2}{\lambda_j^2} - 1 \right) \left(\frac{w_{j'}^2}{\lambda_{j'}^2} - 1 \right) .
\end{aligned} \tag{S-1.33}$$

Note that, since the biochemical factors W_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 (S-1.11).

By using (S-1.30) and the orthonormality of the Hermite polynomials H_1 and H_2, given by (S-1.32), we can obtain the following approximations to the SESI's and JESI's (expressed for the general case of J biochemical factors):

$$
\begin{aligned}
\widehat{\sigma}_j &= \frac{\widehat{V}_j}{\widehat{V}}, \qquad \widehat{\eta}_j = \frac{\widehat{U}_j}{\widehat{V}} \\
\widehat{V}_j &= \alpha^2_{j,1} + \alpha^2_{j,2} \\
\widehat{V}_{jj'} &= \alpha^2_{jj',1} + \alpha^2_{jj',2} + \alpha^2_{jj',3} + \alpha^2_{jj',4} \\
\widehat{U}_j &= \sum_{j'=1, j' \neq j}^{J} \widehat{V}_{jj'} \\
\widehat{V} &= \sum_{j=1}^{J} \widehat{V}_j + \sum_{j=1}^{J-1} \sum_{j'=j+1}^{J} \widehat{V}_{jj'}
\end{aligned}
\tag{S-1.34}
$$

We respectively refer to $\widehat{\sigma}_j$ and $\widehat{\eta}_j$, given by (S-1.34), as the SESI's and JESI's obtained by *Orthonormal Hermite Approximation* (OHA) of the ANOVA-HDMR.

3 Numerical implementation

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

3.1 TSMR

Approximation of the SESI's and JESI's by means of (S-1.16) requires evaluation of the first- and second-order partial derivatives of the response function $R(\boldsymbol{w})$ at $\boldsymbol{w} = \mathbf{0}$, given by

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

Unfortunately, accurate evaluation of these derivatives is not an easy task [9]. 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 [10]. 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 [10].

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

$$d_j \simeq \frac{R(\Delta \boldsymbol{e}_j) - R(-\Delta \boldsymbol{e}_j)}{2\Delta}\,, \tag{S-1.35}$$

for a sufficiently small differential step size $\Delta > 0$, where $\boldsymbol{e}_j$ 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 $R(\boldsymbol{w})$ with respect to w_j and $w_{j'}$ at $\mathbf{0}$:

$$d_{jj'} \simeq \frac{R(\Delta \boldsymbol{e}_j + \Delta \boldsymbol{e}_{j'}) - R(-\Delta \boldsymbol{e}_j + \Delta \boldsymbol{e}_{j'}) - R(\Delta \boldsymbol{e}_j - \Delta \boldsymbol{e}_{j'}) + R(-\Delta \boldsymbol{e}_j - \Delta \boldsymbol{e}_{j'})}{4\Delta^2}\,. \tag{S-1.36}$$

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.

3.2 FD-HDMR

3.2.1 Polynomial approximation

The approximation of the SESI's and JESI's by means of (S-1.20) requires knowledge of the values of the α parameters associated with the polynomial approximation of the basis functions r, given by (S-1.19). This can be done by polynomial regression [11], as we explain next.

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

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

and

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

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 (S-1.3) to evaluate the basis functions r at a set $\{w_j(q), q \in S, j \in J\}$ of prespecified factor values around zero. Then, the least-squares error estimates $\widehat{\alpha}_{j,1}, \widehat{\alpha}_{j,2}$ of the parameters $\alpha_{j,1}, \alpha_{j,2}$ associated with the basis function $r_j(w_j)$ are given by [11]:

$$\widehat{\boldsymbol{\alpha}}_j = (\mathbb{W}_j^T\, \mathbb{W}_j)^{-1} \mathbb{W}_j^T\, \boldsymbol{r}_j\,,$$

where

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

provided that the matrix $\mathbb{W}_j^T\mathbb{W}_j$ is invertible (which is always true if no column of the $\mathbb{W}_j$ matrix is a linear combination of the other columns). On the other hand, the least-squares 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 $\widehat{r}_{jj'}(w_j, w_{j'})$ are given by [11]:

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

where

$$\widehat{\boldsymbol{\alpha}}_{jj'} := \begin{bmatrix} \widehat{\alpha}_{jj',1} \\ \widehat{\alpha}_{jj',2} \\ \widehat{\alpha}_{jj',3} \\ \widehat{\alpha}_{jj',4} \end{bmatrix}_{4\times 1} \qquad \boldsymbol{r}_{jj'} := \begin{bmatrix} r_{jj'}(w_j(1), w_{j'}(1)) \\ \vdots \\ r_{jj'}(w_j(1), w_{j'}(S)) \\ r_{jj'}(w_j(2), w_{j'}(1)) \\ \vdots \\ r_{jj'}(w_j(2), w_{j'}(S)) \\ \vdots \\ r_{jj'}(w_j(S), w_{j'}(1)) \\ \vdots \\ r_{jj'}(w_j(S), w_{j'}(S)) \end{bmatrix}_{S^2\times 1}$$

$$\mathbb{W}_{jj'} := \begin{bmatrix} w_j(1)w_{j'}(1) & w_j^2(1)w_{j'}(1) & w_j(1)w_{j'}^2(1) & w_j^2(1)w_{j'}^2(1) \\ \vdots & \vdots & \vdots & \vdots \\ w_j(1)w_{j'}(S) & w_j^2(1)w_{j'}(S) & w_j(1)w_{j'}^2(S) & w_j^2(1)w_{j'}^2(S) \\ w_j(2)w_{j'}(1) & w_j^2(2)w_{j'}(1) & w_j(2)w_{j'}^2(1) & w_j^2(2)w_{j'}^2(1) \\ \vdots & \vdots & \vdots & \vdots \\ w_j(2)w_{j'}(S) & w_j^2(2)w_{j'}(S) & w_j(2)w_{j'}^2(S) & w_j^2(2)w_{j'}^2(S) \\ w_j(S)w_{j'}(1) & w_j^2(S)w_{j'}(1) & w_j(S)w_{j'}^2(1) & w_j^2(S)w_{j'}^2(1) \\ \vdots & \vdots & \vdots & \vdots \\ w_j(S)w_{j'}(S) & w_j^2(S)w_{j'}(S) & w_j(S)w_{j'}^2(S) & w_j^2(S)w_{j'}^2(S) \end{bmatrix}_{S^2\times 4},$$

provided that the matrix $\mathbb{W}_{jj'}^T \mathbb{W}_{jj'}$ 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.

3.2.2 Gauss-Hermite Integration

It is clear from (S-1.24) and (S-1.25) that evaluation of the SESI's and JESI's by (S-1.27) requires calculation of the expectations $\mathrm{E}[\psi_m(W_m)]$, $\mathrm{E}[\psi_{jm'}(w_j, W_m)]$, $\mathrm{E}[\psi_{mj}(W_m, w_j)]$, $\mathrm{E}[\psi_{mm'}(W_m, W_{m'})]$, $\mathrm{E}[e_j^2(w_j)]$, and $\mathrm{E}[e_{jj'}^2(w_j, w_{j'})]$ with respect to Gaussian distributions. We can evaluate these expectations by using Gauss-Hermite integration [12], as we explain next.

Let us consider the one-dimensional expectation:

$$E_1 = \mathrm{E}[\psi_1(W_1)] = \frac{1}{\lambda\sqrt{2\pi}} \int_{-\infty}^{\infty} \psi_1(w_1) e^{-w_1^2/2\lambda^2} dw_1.$$

If we set $w_1 = \sqrt{2}\lambda u_1$, then

$$E_1 = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \psi_1(\sqrt{2}\lambda u_1) e^{-u_1^2} du_1.$$

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

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

where Q is the order of the approximation and a_q, ω_q are appropriately chosen abscissas and weights, respectively [12].

Likewise, by setting $w_1 = \sqrt{2}\lambda_1 u_1$ and $w_2 = \sqrt{2}\lambda_2 u_2$, we can write the two-dimensional expectation

$$E_2 = \mathrm{E}[\psi_2(W_1, W_2)] = \frac{1}{2\pi\lambda_1\lambda_2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi_2(w_1, w_2) e^{-w_1^2/2\lambda_1^2} e^{-w_2^2/2\lambda_2^2} dw_1 dw_2$$

in the form

$$E_2 = \frac{1}{\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi_2(\sqrt{2}\lambda_1 u_1, \sqrt{2}\lambda_2 u_2) e^{-u_1^2} e^{-u_2^2} du_1 du_2.$$

A two-step (first for u_1 and then for u_2) application of one-dimensional Gauss-Hermite integration results in the following approximation of E_2:

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

It turns out that calculation of the expectations required by (S-1.27) 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 number Q of points used by Gauss-Hermite integration.

3.3 ANOVA-HDMR

Approximating the sensitivity indices σ_j and η_j by (S-1.34) requires evaluation of the parameters α so that the functions $\widehat{\rho}$, given by (S-1.33), result in a sufficiently good approximation of the response function R by $\widehat{R}$, given by (S-1.31). Our problem here is to estimate the parameters α, so that

$$\rho_j(w_j) = \widehat{\rho}_j(w_j) + \epsilon_j = \alpha_{j,1}\frac{w_j}{\lambda_j} + \frac{\alpha_{j,2}}{\sqrt{2}}\left(\frac{w_j^2}{\lambda_j^2}-1\right) + \epsilon_j, \tag{S-1.37}$$

and

$$\begin{aligned}\rho_{jj'}(w_j,w_{j'}) &= \widehat{\rho}_{jj'}(w_j,w_{j'}) + \epsilon_{jj'} \\ &= \alpha_{jj',1}\frac{w_j w_{j'}}{\lambda_j\lambda_{j'}} + \frac{\alpha_{jj',2}}{\sqrt{2}}\left(\frac{w_j^2}{\lambda_j^2}-1\right)\frac{w_{j'}}{\lambda_{j'}} + \frac{\alpha_{jj',3}}{\sqrt{2}}\frac{w_j}{\lambda_j}\left(\frac{w_{j'}^2}{\lambda_{j'}^2}-1\right) \\ &\quad + \frac{\alpha_{jj',4}}{2}\left(\frac{w_j^2}{\lambda_j^2}-1\right)\left(\frac{w_{j'}^2}{\lambda_{j'}^2}-1\right) + \epsilon_{jj'}\,,\end{aligned} \tag{S-1.38}$$

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 (S-1.37), note that

$$\begin{aligned}\alpha_{j,1} &= \int_{-\infty}^{\infty}\frac{w_j}{\lambda_j}\rho_j(w_j)G_j(w_j)dw_j \\ \alpha_{j,2} &= \frac{1}{\sqrt{2}}\int_{-\infty}^{\infty}\left(\frac{w_j^2}{\lambda_j^2}-1\right)\rho_j(w_j)G_j(w_j)dw_j\,,\end{aligned} \tag{S-1.39}$$

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

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

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 (S-1.38), we have that

$$\begin{aligned}
\alpha_{jj',1} &= \int_{-\infty}^{\infty} \frac{w_j}{\lambda_j}\frac{w_{j'}}{\lambda_{j'}}\,\rho_{jj'}(w_j,w_{j'})G_j(w_j)G_{j'}(w_{j'})dw_j dw_{j'} \\
\alpha_{jj',2} &= \frac{1}{\sqrt{2}}\int_{-\infty}^{\infty} \left(\frac{w_j^2}{\lambda_j^2}-1\right)\frac{w_{j'}}{\lambda_{j'}}\,\rho_{jj'}(w_j,w_{j'})G_j(w_j)G_{j'}(w_{j'})dw_j dw_{j'} \\
\alpha_{jj',3} &= \frac{1}{\sqrt{2}}\int_{-\infty}^{\infty} \frac{w_j}{\lambda_j}\left(\frac{w_{j'}^2}{\lambda_{j'}^2}-1\right)\,\rho_{jj'}(w_j,w_{j'})G_j(w_j)G_{j'}(w_{j'})dw_j dw_{j'} \\
\alpha_{jj',4} &= \frac{1}{2}\int_{-\infty}^{\infty} \left(\frac{w_j^2}{\lambda_j^2}-1\right)\left(\frac{w_{j'}^2}{\lambda_{j'}^2}-1\right)\,\rho_{jj'}(w_j,w_{j'})G_j(w_j)G_{j'}(w_{j'})dw_j dw_{j'}.
\end{aligned} \tag{S-1.40}$$

Finally,

$$\begin{aligned}
\alpha_{j,1} &= \mathrm{E}\left[\frac{W_j}{\lambda_j}R(\boldsymbol{W})\right] \\
\alpha_{j,2} &= \frac{1}{\sqrt{2}}\,\mathrm{E}\left[\left(\frac{W_j^2}{\lambda_j^2}-1\right)R(\boldsymbol{W})\right] \\
\alpha_{jj',1} &= \mathrm{E}\left[\frac{W_j}{\lambda_j}\frac{W_{j'}}{\lambda_{j'}}R(\boldsymbol{W})\right] \\
\alpha_{jj',2} &= \frac{1}{\sqrt{2}}\,\mathrm{E}\left[\left(\frac{W_j^2}{\lambda_j^2}-1\right)\frac{W_{j'}}{\lambda_{j'}}R(\boldsymbol{W})\right] \\
\alpha_{jj',3} &= \frac{1}{\sqrt{2}}\,\mathrm{E}\left[\frac{W_j}{\lambda_j}\left(\frac{W_{j'}^2}{\lambda_{j'}^2}-1\right)R(\boldsymbol{W})\right] \\
\alpha_{jj',4} &= \frac{1}{2}\,\mathrm{E}\left[\left(\frac{W_j^2}{\lambda_j^2}-1\right)\left(\frac{W_{j'}^2}{\lambda_{j'}^2}-1\right)R(\boldsymbol{W})\right],
\end{aligned} \tag{S-1.41}$$

by virtue of (S-1.10), (S-1.39), and (S-1.40).

As a consequence of the previous analysis, to determine the parameters α, we need to evaluate the expectations in (S-1.41). 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 [13, 14]. In particular, we can generate L Latin hypercube Gaussian samples $\boldsymbol{w}^{(l)} = \{w_1^{(l)}, w_2^{(l)}, \ldots, w_J^{(l)}\}$,

$l = 1, 2, \ldots, L$, evaluate the responses $R(\boldsymbol{w}^{(l)})$, for $l = 1, 2, \ldots, L$, and set

$$\alpha_{j,1} \simeq \widehat{\alpha}_{j,1} := \frac{1}{L} \sum_{l=1}^{L} \frac{w_j^{(l)}}{\lambda_j} R(\boldsymbol{w}^{(l)})$$

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

$$\alpha_{jj',1} \simeq \widehat{\alpha}_{jj',1} := \frac{1}{L} \sum_{l=1}^{L} \frac{w_j^{(l)}}{\lambda_j} \frac{w_{j'}^{(l)}}{\lambda_{j'}} R(\boldsymbol{w}^{(l)})$$

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

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

$$\alpha_{jj',4} \simeq \widehat{\alpha}_{jj',4} := \frac{1}{2} \frac{1}{L} \sum_{l=1}^{L} \left(\frac{[w_j^{(l)}]^2}{\lambda_j^2} - 1 \right) \left(\frac{[w_{j'}^{(l)}]^2}{\lambda_{j'}^2} - 1 \right) R(\boldsymbol{w}^{(l)}). \tag{S-1.42}$$

Clearly, implementation of (S-1.42) 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 $\{\boldsymbol{w}^{(l)}, l = 1, 2, \ldots, L\}$ and estimate the α parameters by polynomial regression, as we did in Section 3.2.1. We discuss this approach in the following.

As a consequence of (S-1.31) and (S-1.33), the polynomial regression problem amounts to estimating $\widehat{\rho}_0$ and the parameters α, so that

$$\begin{aligned} R(\boldsymbol{w}) &= \widehat{R}(\boldsymbol{w}) + \epsilon \\ &= \widehat{\rho}_0 + \sum_{j=1}^{J} \alpha_{j,1} \frac{w_j}{\lambda_j} + \frac{\alpha_{j,2}}{\sqrt{2}} \left(\frac{w_j^2}{\lambda_j^2} - 1 \right) \\ &\quad + \sum_{j=1}^{J-1} \sum_{j'=j+1}^{J} \alpha_{jj',1} \frac{w_j w_{j'}}{\lambda_j \lambda_{j'}} + \frac{\alpha_{jj',2}}{\sqrt{2}} \left(\frac{w_j^2}{\lambda_j^2} - 1 \right) \frac{w_{j'}}{\lambda_{j'}} \end{aligned}$$

$$+\sum_{j=1}^{J-1}\sum_{j'=j+1}^{J}\frac{\alpha_{jj',3}}{\sqrt{2}}\frac{w_j}{\lambda_j}\left(\frac{w_{j'}^2}{\lambda_{j'}^2}-1\right)+\frac{\alpha_{jj',4}}{2}\left(\frac{w_j^2}{\lambda_j^2}-1\right)\left(\frac{w_{j'}^2}{\lambda_{j'}^2}-1\right)+\epsilon\,,$$

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-squares error estimate $\widehat{\boldsymbol{\alpha}}$ of the parameters α are given by

$$\widehat{\boldsymbol{\alpha}}=(\mathbb{W}^T\mathbb{W})^{-1}\mathbb{W}^T\boldsymbol{\rho}\,, \tag{S-1.43}$$

where

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

$$\mathbb{W}:=\begin{bmatrix}1 & w_1^{(1)}/\lambda_1 & \cdots & [(w_1^{(1)}/\lambda_1)^2-1]/\sqrt{2} & \cdots & [(w_{J-1}^{(1)}/\lambda_{J-1})^2-1][(w_J^{(1)}/\lambda_J)^2-1]/2\\ 1 & w_1^{(2)}/\lambda_1 & \cdots & [(w_1^{(2)}/\lambda_1)^2-1]/\sqrt{2} & \cdots & [(w_{J-1}^{(2)}/\lambda_{J-1})^2-1][(w_J^{(2)}/\lambda_J)^2-1]/2\\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots\\ 1 & w_1^{(L)}/\lambda_1 & \cdots & [(w_1^{(L)}/\lambda_1)^2-1]/\sqrt{2} & \cdots & [(w_{J-1}^{(L)}/\lambda_{J-1})^2-1][(w_J^{(L)}/\lambda_J)^2-1]/2\end{bmatrix}_{L\times(2J^2+1)}, \tag{S-1.44}$$

provided that the matrix $\mathbb{W}^T\mathbb{W}$ is invertible. Note that calculation of $\widehat{\boldsymbol{\alpha}}$ requires the same number L of system integrations as Monte Carlo estimation by (S-1.42).

It is not difficult to see from (S-1.43) that, if $\widehat{\boldsymbol{\alpha}}_{\text{MC}}$ is the Monte Carlo estimate of $\widehat{\rho}_0$ and of the parameters α, given by [recall (S-1.10)]

$$\widehat{\rho}_0=\rho_0=\mathrm{E}[R(\boldsymbol{W})]\simeq\frac{1}{L}\sum_{l=1}^{L}R(\boldsymbol{w}^{(l)}),$$

and (S-1.42), respectively, then

$$\widehat{\boldsymbol{\alpha}}_{\text{MC}}=\frac{1}{L}\,\mathbb{W}^T\boldsymbol{\rho}=\frac{1}{L}\,\mathbb{W}^T\mathbb{W}\,\widehat{\boldsymbol{\alpha}}\,.$$

Moreover,

$$\lim_{L\to\infty}\frac{1}{L}\mathbb{W}^T\mathbb{W}=\mathbb{I}, \tag{S-1.45}$$

where $\mathbb{I}$ is the identity matrix, by virtue of the biorthonormality conditions given by (S-1.32) and the fact that the Monte Carlo estimate $\sum_{l=1}^{L}f(x^{(l)})/L$ 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 $\widehat{\boldsymbol{\alpha}}_{\text{MC}}$ and the regression estimate $\widehat{\boldsymbol{\alpha}}$ are identical in the limit as the number of Monte Carlo samples grows to infinity. Since $\boldsymbol{\alpha}$ is obtained by minimizing the least-squares error between R and $\widehat{R}$, we expect that the regression estimate $\boldsymbol{\alpha}$ of the parameters α will be more preferable than the Monte Carlo estimate $\widehat{\boldsymbol{\alpha}}_{\text{MC}}$, in the sense that, for a relatively small number of Monte Carlo samples, $\boldsymbol{\alpha}$ may produce a better fit $\widehat{R}$ of the response function R than the one produced by $\widehat{\boldsymbol{\alpha}}_{\text{MC}}$. Finally, note from (S-1.45) that, for a sufficiently large number of Monte Carlo samples, $\mathbb{W}^T\mathbb{W}$ is approximately equal to the identity matrix multiplied by L, which effectively reduces the risk of singularity when evaluating the inverse matrix $(\mathbb{W}^T\mathbb{W})^{-1}$ in (S-1.43). Therefore, if $\mathbb{W}^T\mathbb{W}$ turns out to be singular for a chosen value of L, the user needs to increase L until a nonsingular matrix $\mathbb{W}^T\mathbb{W}$ is obtained.

References

1. Rabitz H, Aliş ÖF, Shorter J, Shim K: **Efficient input-output model representations**. *Comput. Phys. Commun.* 1999, **117**:11–20.
2. Rabitz H, Aliş ÖF: **General foundations of high-dimensional model representations**. *J. Math. Chem.* 1999, **25**:197–233.
3. Li G, Rosenthal C, Rabitz H: **High dimensional model representations**. *J. Phys. Chem. A* 2001, **105**:7765–7777.
4. Sobol' IM: **Theorems and examples on high-dimensional model representation**. *Reliab. Eng. Syst. Safe.* 2003, **79**:187–193.
5. Sobol' IM: **Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates**. *Math. Comput. Simulat.* 2001, **55**:271–280.
6. Chen W, Jin R, Sudjianto A: **Analytical variance-based global sensitivity analysis in simulation-based design under uncertainty**. *J. Mech. Design* 2005, **127**:875–886.
7. Archer GEB, Saltelli A, Sobol IM: **Sensitivity measures, ANOVA-like techniques and the use of bootstrap**. *J. Statist. Comput. Simul.* 1997, **58**:99–120.
8. Owen AB: **Latin supercube sampling for very high-dimensional simulations**. *ACM T. Model. Comput. S.* 1998, **8**:71–102.
9. Cacuci DG: *Sensitivity and Uncertainty Analysis*, *Volume I. Theory*. Boca Raton: Chapman & Hall/CRC 2003.
10. Varma A, Morbidelli M, Wu H: *Parametric Sensitivity in Chemical Systems*. Cambridge, UK: Cambridge University Press 1999.
11. Montgomery DC, Peck EA, Vining GG: *Introduction to Linear Regression Analysis*. New York: John Wiley, 3rd edition 2001.
12. Press WH, Teukolsky SA, Vetterling WT, Flannery BP: *Numerical Recipes: The Art of Scientific Computing*. New York: Cambridge University Press, 3rd edition 2007.
13. Stein M: **Large sample properties of simulations using Latin hypercube sampling**. *Technometrics* 1987, **29**:143–151.
14. McKay MD, Conover WJ, Beckman RJ: **A Comparison of three methods for selecting values of input variables in the analysis of output from a computer code**. *Technometrics* 2000, **42**:55–61.

ADDITIONAL FILE 2

A comparison of approximation techniques for variance-based sensitivity analysis of biochemical reaction systems

MAPK SIGNALING CASCADE MODEL

Hong-Xuan Zhang[1] **and John Goutsias**[*1]

[1] Whitaker Biomedical Engineering Institute, The Johns Hopkins University, Baltimore, MD 21218, USA

[*] Corresponding author

Email: HXZ: hxzhang@jhu.edu, JG: goutsias@jhu.edu

In this document, we list the biochemical reactions associated with the MAPK signaling cascade model we consider in the Main text and provide nominal values for the normalized reaction rate constants (measured in s^{-1}) and the initial molecular concentrations (measured in molecules/cell). We depict this model in Figure 1 of the Main text. We have adopted the data from Schoeberl *et al.* [1], 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 Figure 1 of the Main text compensates for Ras-GTP synthesis which, in reality, is accomplished by a complex epidermal growth factor (EGF)-induced signalling pathway [1]. We have set the reaction 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.* [1], with 50ng/ml EGF.

Reactions

No.	Reaction	Rate Constant (s^{-1})
1	$\emptyset \rightarrow$ Ras-GTP	$\kappa_1 = 3$
	Ras-GTP $\rightarrow \emptyset$	$\kappa_2 = 0$
2	Ras-GTP + Raf $\rightarrow$ Raf-Ras-GTP	$\kappa_3 = 1.6605 \times 10^{-6}$
	Raf-Ras-GTP $\rightarrow$ Raf + Ras-GTP	$\kappa_4 = 5.3 \times 10^{-3}$
3	Raf-Ras-GTP $\rightarrow$ Raf* + Ras-GTP*	$\kappa_5 = 1$
	Raf* + Ras-GTP* $\rightarrow$ Raf-Ras-GTP	$\kappa_6 = 1.1624 \times 10^{-6}$
4	Raf* + Pho1 $\rightarrow$ Raf*-Pho1	$\kappa_7 = 1.1790 \times 10^{-4}$
	Raf*-Pho1 $\rightarrow$ Raf* + Pho1	$\kappa_8 = 0.2$
5	Raf*-Pho1 $\rightarrow$ Raf + Pho1	$\kappa_9 = 1$
	Raf + Pho1 $\rightarrow$ Raf*-Pho1	$\kappa_{10} = 0$
6	MEK + Raf* $\rightarrow$ MEK-Raf*	$\kappa_{11} = 1.9428 \times 10^{-5}$
	MEK-Raf* $\rightarrow$ MEK + Raf*	$\kappa_{12} = 3.3 \times 10^{-2}$
7	MEK-Raf* $\rightarrow$ MEK-P + Raf*	$\kappa_{13} = 3.5$
	MEK-P + Raf* $\rightarrow$ MEK-Raf*	$\kappa_{14} = 0$
8	MEK-P + Raf* $\rightarrow$ MEK-P-Raf*	$\kappa_{15} = 1.9428 \times 10^{-5}$
	MEK-P-Raf* $\rightarrow$ MEK-P + Raf*	$\kappa_{16} = 3.3 \times 10^{-2}$
9	MEK-P-Raf* $\rightarrow$ MEK-PP + Raf*	$\kappa_{17} = 2.9$
	MEK-PP + Raf* $\rightarrow$ MEK-P-Raf*	$\kappa_{18} = 0$
10	MEK-PP + Pho2 $\rightarrow$ MEK-PP-Pho2	$\kappa_{19} = 2.3746 \times 10^{-5}$
	MEK-PP-Pho2 $\rightarrow$ MEK-PP + Pho2	$\kappa_{20} = 0.8$
11	MEK-PP-Pho2 $\rightarrow$ MEK-P + Pho2	$\kappa_{21} = 5.8 \times 10^{-2}$
	MEK-P + Pho2 $\rightarrow$ MEK-PP-Pho2	$\kappa_{22} = 0$
12	MEK-P + Pho2 $\rightarrow$ MEK-P-Pho2	$\kappa_{23} = 4.4835 \times 10^{-7}$
	MEK-P-Pho2 $\rightarrow$ MEK-P + Pho2	$\kappa_{24} = 0.5$
13	MEK-P-Pho2 $\rightarrow$ MEK + Pho2	$\kappa_{25} = 5.8 \times 10^{-2}$
	MEK + Pho2 $\rightarrow$ MEK-P-Pho2	$\kappa_{26} = 0$
14	ERK + MEK-PP $\rightarrow$ ERK-MEK-PP	$\kappa_{27} = 8.8673 \times 10^{-5}$
	ERK-MEK-PP $\rightarrow$ ERK + MEK-PP	$\kappa_{28} = 1.833 \times 10^{-2}$
15	ERK-MEK-PP $\rightarrow$ ERK-P + MEK-PP	$\kappa_{29} = 16$
	ERK-P + MEK-PP $\rightarrow$ ERK-MEK-PP	$\kappa_{30} = 0$
16	ERK-P + MEK-PP $\rightarrow$ ERK-P-MEK-PP	$\kappa_{31} = 8.8673 \times 10^{-5}$
	ERK-P-MEK-PP $\rightarrow$ ERK-P + MEK-PP	$\kappa_{32} = 1.833 \times 10^{-2}$
17	ERK-P-MEK-PP $\rightarrow$ ERK-PP + MEK-PP	$\kappa_{33} = 5.7$
	ERK-PP + MEK-PP $\rightarrow$ ERK-P-MEK-PP	$\kappa_{34} = 0$
18	ERK-PP + Pho3 $\rightarrow$ ERK-PP-Pho3	$\kappa_{35} = 2.3414 \times 10^{-5}$
	ERK-PP-Pho3 $\rightarrow$ ERK-PP + Pho3	$\kappa_{36} = 0.6$
19	ERK-PP-Pho3 $\rightarrow$ ERK-P + Pho3	$\kappa_{37} = 0.246$
	ERK-P + Pho3 $\rightarrow$ ERK-PP-Pho3	$\kappa_{38} = 0$
20	ERK-P + Pho3 $\rightarrow$ ERK-P-Pho3	$\kappa_{39} = 8.3027 \times 10^{-6}$
	ERK-P-Pho3 $\rightarrow$ ERK-P + Pho3	$\kappa_{40} = 0.5$
21	ERK + Pho3 $\rightarrow$ ERK-P-Pho3	$\kappa_{41} = 0$
	ERK-P-Pho3 $\rightarrow$ ERK + Pho3	$\kappa_{42} = 0.246$

Initial Concentrations

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

References

1. Schoeberl B, Eichler-Jonsson C, Gilles ED, Müller G: **Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors**. *Nat. Biotechnol.* 2002, **20**:370–375.

ADDITIONAL FILE 3

A comparison of approximation techniques for variance-based sensitivity analysis of biochemical reaction systems

SOSA-BASED SENSITIVITY ANALYSIS RESULTS

Hong-Xuan Zhang[1] **and John Goutsias**[*1]

[1] Whitaker Biomedical Engineering Institute, The Johns Hopkins University, Baltimore, MD 21218, USA

[*] Corresponding author

Email: HXZ: hxzhang@jhu.edu, JG: goutsias@jhu.edu

In this document, 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 the Main text 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. 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 the Main text, 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 *noninfluential*.

Table S-3.1. SOSA-based sensitivity analysis results for the *duration* of ERK-PP activity.

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 S-3.2. SOSA-based sensitivity analysis results for the *integrated response* of ERK-PP activity.

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 S-3.3. SOSA-based sensitivity analysis results for the *strength* of ERK-PP activity.

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