AGU PUBLICATIONS

Water Resources Research

REVIEW ARTICLE
10.1002/2014WR016527

**Key Points:**
- No unique definition exists for "global sensitivity" across the problem space
- Existing approaches to sensitivity analysis are limited in consistency and utility
- We identify important properties of response surfaces that relate to sensitivity

**Correspondence to:**
S. Razavi,
saman.razavi@usask.ca

**Citation:**
Razavi, S., and H. V. Gupta (2015), What do we mean by sensitivity analysis? The need for comprehensive characterization of "global" sensitivity in Earth and Environmental systems models, *Water Resour. Res.*, *51*, 3070–3092, doi:10.1002/2014WR016527.

Received 8 OCT 2014
Accepted 17 MAR 2015
Accepted article online 23 MAR 2015
Published online 23 MAY 2015



# What do we mean by sensitivity analysis? The need for comprehensive characterization of "global" sensitivity in Earth and Environmental systems models

**Saman Razavi[1] and Hoshin V. Gupta[2]**

[1]Global Institute for Water Security, School of Environment and Sustainability, Department of Civil and Geological Engineering, University of Saskatchewan, Saskatoon, Saskatchewan, Canada, [2]Department of Hydrology and Water Resources, University of Arizona, Tucson, Arizona, USA

**Abstract** Sensitivity analysis is an essential paradigm in Earth and Environmental Systems modeling. However, the term "sensitivity" has a clear definition, based in partial derivatives, only when specified locally around a particular point (e.g., optimal solution) in the problem space. Accordingly, no unique definition exists for "global sensitivity" across the problem space, when considering one or more model responses to different factors such as model parameters or forcings. A variety of approaches have been proposed for global sensitivity analysis, based on different philosophies and theories, and each of these formally characterizes a different "intuitive" understanding of sensitivity. These approaches focus on different properties of the model response at a fundamental level and may therefore lead to different (even conflicting) conclusions about the underlying sensitivities. Here we revisit the theoretical basis for sensitivity analysis, summarize and critically evaluate existing approaches in the literature, and demonstrate their flaws and shortcomings through conceptual examples. We also demonstrate the difficulty involved in interpreting "global" interaction effects, which may undermine the value of existing interpretive approaches. With this background, we identify several important properties of response surfaces that are associated with the understanding and interpretation of sensitivities in the context of Earth and Environmental System models. Finally, we highlight the need for a new, comprehensive framework for sensitivity analysis that effectively characterizes all of the important sensitivity-related properties of model response surfaces.

## 1. Introduction

Sensitivity analysis (SA) is an important paradigm in the context of model development and application. There exist a variety of approaches toward SA that formally describe different "intuitive" understandings of the sensitivity of one or more model responses to different factors such as model parameters or forcings. Further, the objectives of SA can vary with application, and a survey of the literature reveals that it has been used to explore a variety of aspects and questions pertaining to model development and application. For example, some of the different (in cases complementary) objectives of SA include:

a. *Assessment of Similarity*: Diagnostic testing and evaluation of similarities between the functioning of the model and the underlying system, so as to assess fidelity of the model structure and conceptualization [e.g., *Clark et al*., 2011; *Gupta et al*., 2008; *Saltelli et al*., 2004; *Spear and Hornberger*, 1980].

b. *Factor Importance and Function*: Identification, prioritization, and screening of the factors that are more influential and contribute most significantly to variability and other characteristics of model/system response [e.g., *Muleta and Nicklow*, 2005; *Ruano et al*., 2012].

c. *Regions of Sensitivity*: Location and characterization of regions in the factor space where the model/system presents the highest variability in response to variations in the factors [e.g., *Rakovec et al*., 2014]. This is instrumental in model calibration.

d. *Factor Interdependence*: Investigation of the nature and strength of interactions between the factors, and the degree to which factors intensify, cancel, or compensate for the effects of each other [e.g., *Lidén and Harlin*, 2000; *Nossent et al*., 2011; *Ratto et al*., 2001].

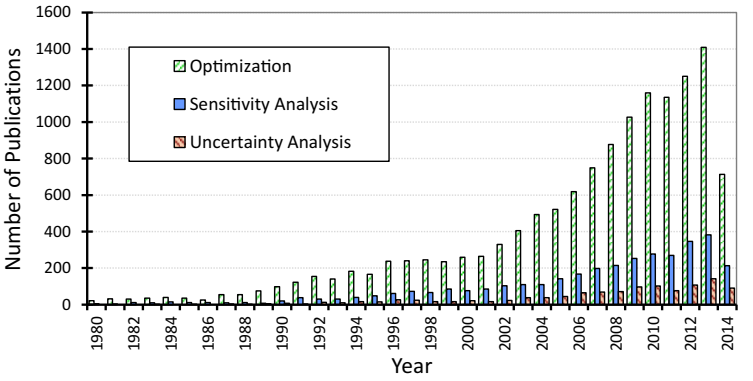


**Figure 1.** History of the number of publications with the topic "Sensitivity Analysis" within the research area "Water Resources" based on a search in Thomson Reuters Web of Science (August 2014). Also shown are results within the same domain for searches on "Optimization" or "Optimisation," and on "Uncertainty Analysis" or "Uncertainty Quantification."

e. *Factor and Model Reduction:* Identification of the noninfluential factors and/or insensitive (possibly redundant) components of model structure, usually so they can be constrained or removed so as to simplify the model/analysis [e.g., *Degenring et al*., 2004; *Nossent et al*., 2011; *van Griensven et al*., 2006].

f. *Uncertainty Apportionment:* Quantitative attribution of the uncertainty in model response to different factors (sources of uncertainty), with the goal of identifying where best to focus efforts at improved factor characterization so as to achieve reductions in total uncertainty [e.g., *Chu-Agor et al*., 2011; *Crosetto et al*., 2000].

Over the past two decades the interest in SA has increased considerably. A search (Figure 1) of Thomson Reuters Web of Science for the topic "sensitivity analysis" shows the importance of SA in relation to the related important topics of optimization and uncertainty in the field of "water resources." Further, these publications from the field of water resources constitute about 7%, 3%, and 1% respectively of the corresponding publications on Optimization, Sensitivity Analysis, and Uncertainty Analysis found in the full body of literature from all fields (according to Thomson Reuters Web of Science).

The significance of SA and the associated challenges in the context of Earth and Environmental Systems modeling cannot be understated. Such models are rapidly becoming increasingly more complex and computationally intensive, growing in dimensionality (both process and parameter), as they progressively and more rigorously reflect our growing understanding (or hypotheses) about the underlying real-world systems they are constructed to represent. However, while SA has become a critical tool in the development and application of such models, its widespread applicability can be inhibited by computational expense. It seems clear, therefore, that the development of strategies for SA that are both effective and efficient is of paramount importance.

### 1.1. Challenges

Despite significant advancements in the development and application of SA techniques, two issues continue to pose major challenges:

1. Ambiguous Definition of Sensitivity – A variety of methods for SA have been presented in the literature, based in different philosophies and theoretical definitions of sensitivity. The absence of a *unique* definition for sensitivity can result in different, even conflicting, conclusions regarding the underlying sensitivities for a given problem.
2. Computational Cost – The cost of carrying out SA varies significantly for different methods, where cost (or "efficiency") is commonly assessed in terms of the number of samples (model simulation runs) required for the method to generate statistically robust and stable results. This cost can be large, even excessive, for high-dimensional problems and/or computationally intensive models.

Further, because different SA methods focus on different characteristic properties of the model response surface, their applicability and suitability in regards to the different objectives outlined above vary. Therefore, a user must carefully consider at least three questions when selecting and applying a method of SA to a given problem:

1. What is the objective of performing sensitivity analysis?
2. What is the intended definition for (meaning of) sensitivity in the current context?

3. What is the computational budget available for sensitivity analysis?

Among these, the second question is perhaps the most important. It is clearly nontrivial, and will influence the specification of the metric(s) to be used for evaluation of model behavior or performance [*Gupta et al*., 1998, 2008, 2009; *Rosolem et al*., 2012]. In the literature regarding SA, the significance of this issue has been largely ignored.

As outlined in section 1.2, the primary purpose of this paper is to address the first major challenge mentioned above. The second challenge (i.e., computational cost/efficiency) is only barely touched upon in this paper, and while additional research to develop more efficient SA approaches is clearly necessary, one common way to circumvent the limiting computational costs associated with SA for computationally intensive Earth and Environmental System Models (EESMs) is to utilize "surrogate modeling" approaches [*Razavi et al*., 2012b], where the original model is replaced by a fast-to-run surrogate model (see *Razavi et al*. [2012a] for a review of such approaches and a list of studies utilizing surrogates for sensitivity analysis of EESMs).

### 1.2. Objectives and Scope

The first part of this paper summarizes concepts and philosophies about SA in a range of fields. It then seeks to establish a firmer theoretical and practical foundation for SA in the context of EESMs. Simple conceptual examples are used to examine and illustrate the theoretical basis for SA, and to critically evaluate existing approaches and discuss and demonstrate their strengths and shortcomings. Based on insights gained through this analysis, we identify the different important properties of a response surface that relate to an intuitive understanding of SA, thereby addressing the first major challenge mentioned above (i.e., the *definition* of sensitivity). We further discuss the difficulty of characterizing the interaction effect between factors and how such interactions are quantified and interpreted via different approaches.

For purposes of this study, we will assume that the user has selected a metric that is both meaningful and clearly reflective of the objectives of the analysis, and focus primarily on the aspect of properly extracting information regarding the behavior of the selected metric in the factor space. In a subsequent paper, we will examine the issue of metric selection in the specific context of SA. Further, we restrict our current attention to the univariate case, where model response is characterized by a single metric. Model and case-study-specific problems related to SA (e.g., the impact of the length of an observation period in a study area on the assessment of sensitivity of a rainfall-runoff model) are not within the scope of this paper. Of course, such problems are also quite important and need to be analyzed on a case-by-case basis; for example *Shin et al*. [2013] provide such an analysis for conceptual rainfall-runoff models.

## 2. The Conceptual Basis for Sensitivity Analysis

### 2.1. Characterization of Response Surfaces

A "response surface" of a model refers to a line, plane, or hyper-plane that represents the variation of a target response of the model (a state or output variable, or a performance metric) with respect to variations in one or more factors of interest. Factors are features of the model (such as model parameters or forcings, etc.), which may vary on continuous, discrete, or hybrid domains that define the "factor space." Figure 2 shows a simple hypothetical example of a two-factor ($x_1$ and $x_2$) response surface, such as might be found in a typical textbook; for example, the quantity $y$ can be a model performance metric (such as a Likelihood function) and the factors can correspond to two of the model parameters. Such a response surface expresses important information regarding the underlying characteristics of the model, including about sensitivities and factor interactions. However, effective characterization of the properties of a model response surface is not trivial, especially when the surface is nonconvex and/or multimodal, and when the computational cost of obtaining a representative sample of points across the factor space is large.

Optimization is perhaps the most common approach used to (partially) characterize the response surface of a model performance metric. It does so by attempting to (efficiently) locate the point on the response surface that maximizes the goodness-of-fit between the input-state-output behaviors of the model and the real system [*Duan et al*., 1992]. More comprehensive characterization of response surfaces can be achieved by an uncertainty analysis based in Bayesian inference, where instead of seeking a single (optimal) point in the factor space, the approach identifies and characterizes the so-called "high probability region" of the factor space where model performance exceeds some specified threshold value [*Vrugt et al*., 2003]. However,

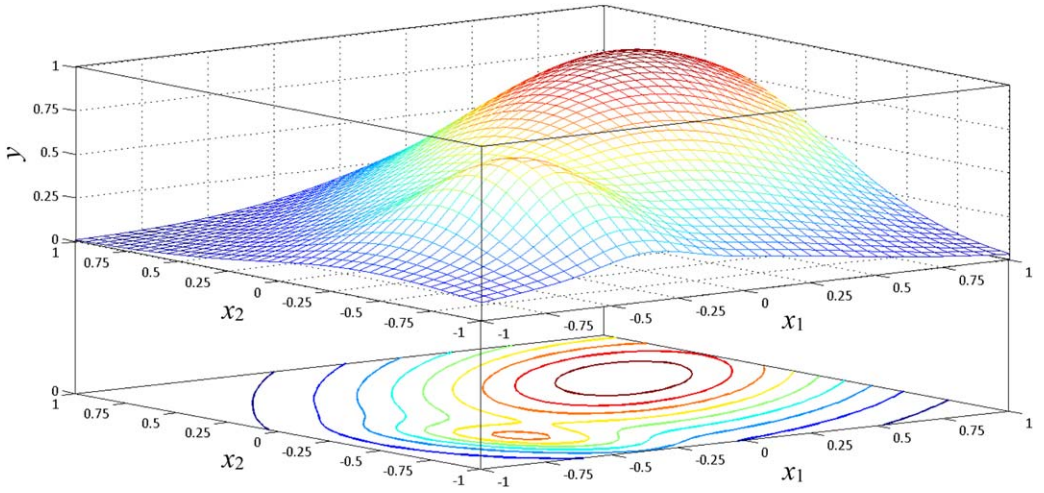


**Figure 2.** Hypothetical response surface representing a model response, $y$, with respect to two factors, $x_1$ and $x_2$.

the most comprehensive paradigm is SA, where "global" sensitivity analysis techniques seek to characterize the properties of a response surface across the entire factor space.

Note that depicting the full shape of the response surfaces of EESMs with more than a few factors, and their full characterization, is typically not possible. Only in the *idealized* case where we had a model with $n$ noninteracting factors (all factor effects were additive), the response surface of the model could be possibly depicted (e.g., through $n$ cross sections of the response surface along the $n$ dimensions at any point in the factor space). Therefore, efforts to do so have focused on simplified models with small numbers of factors (mainly model parameters) through generating contour/3-D plots of pairs of factors [see e.g., *Duan et al*., 1992; *Kuczera and Parent*, 1998]. Depicting and understanding such pseudo response surfaces of EESMs have been extremely useful to explain model behavior and improve our capability in system identification and model development.

To motivate the discussion that follows, and to show that our discussion is not purely academic or hypothetical, Figure 3 shows several two-parameter response surfaces from a 45-parameter coupled land surface scheme-hydrology model (the MESH model, introduced by *Pietroniro et al*., [2007]), illustrating a variety of response surface forms, from relatively simple and smooth to highly complex. The responses of three different model performance metrics to parameter variations are shown; the Nash Sutcliffe efficiency (NS), the Nash Sutcliffe efficiency computed on the log transformed model output (NS_log), and the volume bias (VBIAS). These plots help to illustrate three important points: (1) That the scale at which the response surface is analyzed is important; (2) That the choice of target model response (here the performance metric) affects the form of the response surface; and (3) That the form of response surface can be different in different regions of the parameter space.

In regards to the first point (scale), Figure 3b demonstrates the occurrence of small-scale roughness (likely due to numerical artefacts) that may not be easily visible at larger scales (Figure 3a). In regards to the second point (target model response), Figures 3a, 3c, and 3d show that the NS_log, NS and VBIAS metric response surfaces defined on the same parameter space are significantly different. Here we see that NS_log and VBIAS show strong sensitivity to variations in parameter SAND11 (which partly controls the storage in the catchment) but NS shows much less sensitivity. In contrast, NS is quite sensitive to parameter WF_R2 (which controls the timing of flows), but VBIAS shows no sensitivity at all. Similar behaviors can be seen in Figures 3e–3g. In regards to the third point (region of the parameter space), Figures 3f and 3h show significant differences in the NS response surfaces for parameters SAND22 and SAND12 when the remaining 43 parameters are fixed at different values; in Figure 3f, NS is more sensitive to variations in SAND22 than to variations in SAND12, while in Figure 3h the situation is reversed. This illustrates the role of high-order parameter interactions in EESM models, where any local/subspace sensitivity analysis can provide misleading information about global sensitivity.

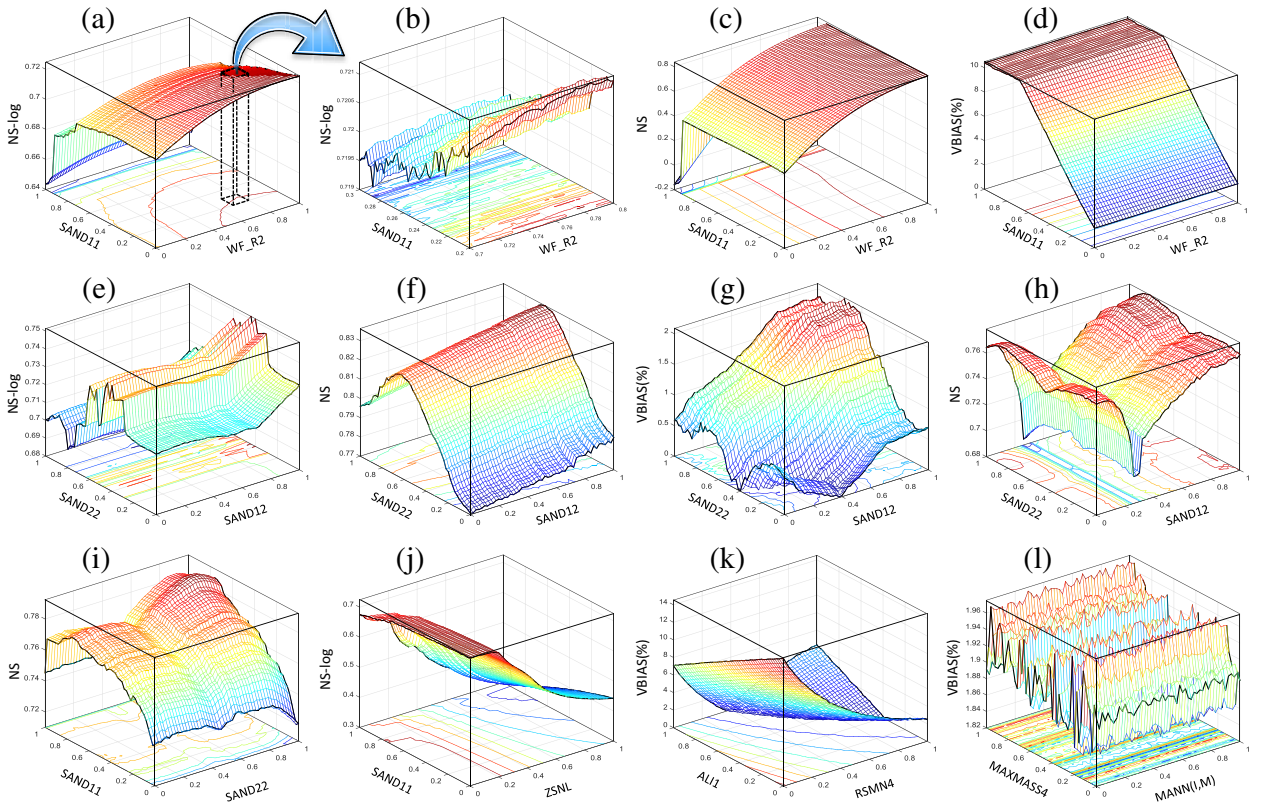


**Figure 3.** Metric response surfaces in two-parameter spaces (scaled on the range 0–1) for a 45-parameter coupled land surface scheme-hydrology model (with other parameters fixed at nominal values). The vertical axes are Nash-Sutcliffe efficiency between observed and simulated flows (NS), Nash-Sutcliffe efficiency on log-transformed values of flows (NS_log), and volume bias (VBIAS). The model parameters include SAND11 (% of sand in soil layer 1 of grouped response unit 1 or GRU1), WF_RF (river roughness factor), SAND22 (% of sand in soil layer 2 of GRU2), SAND12 (% of sand in soil layer 2 of GRU1), ZSNL (limiting snow depth below which coverage is $< 100\%$), ALI1 (near infrared Needle Leaf albedo), RSMN4 (minimum stomatal resistance of Grass), MAXMASS4 (standing biomass density of Grass), and MANN(I,M) (the "Manning's n" constant). In Figures 3a–3d, all other parameters are fixed at the same set of values. In Figures 3e–3g, all other parameters are fixed at the same set of values, but in Figure 3h, they are fixed at a different set of values.

The next section discusses the fundamental concepts of SA, and traces the transition from basic principles (i.e., the definition of "local" sensitivity) into the more advanced concepts required for a "global" characterization of sensitivity.

## 2.2. Local (Point-Based) Sensitivity Analysis (LSA)

### 2.2.1. LSA of Model Response

The "sensitivity" of a model response $y$ to a factor $x$ is defined as the "rate of change (slope)" $\frac{\partial y}{\partial x}$ of the response $y$ in the direction of increasing values of the factor $x$. Suppose that the response of the model is represented by a function $f$ as:

$$y = f(x_1, \ \ldots, \ x_n) \tag{1}$$

where $x_1, \ \ldots, \ x_n$ are factors of interest varying within a factor space defined by the $n$-dimensional hypercube bounded between $x_1^{min}, \ \ldots, \ x_n^{min}$ and $x_1^{max}, \ \ldots, \ x_n^{max}$. The rate of change, $s_i$, of response $y$ with respect to factor $x_i$ ($1 \leq i \leq n$) can be evaluated *at a specific base point* $(x_1^*, \ \ldots, \ x_n^*)$ in the factor space, as the partial derivative of $y$ with respect to $x_i$ at that location:

$$s_i = \left(\frac{\partial y}{\partial x_i}\right)_{(x_1^*, \ \ldots, \ x_n^*)} \tag{2}$$

$s_i$ is sometimes referred to as the "sensitivity coefficient," and characterizes the independent effect of factor $x_i$, when all other factors are held constant.

The covarying effects of multiple factors (i.e., the sensitivity of the model response to "interactions" among the factors) are defined using higher-order partial derivatives. For example, the two-factor interaction effect

of $x_i$ and $x_j$ (where $i \neq j$ and $1 \leq i, j \leq n$) on the model response is represented using a second-order partial derivative, and is interpreted as the rate of change of slope in the direction of $x_i$ as we move in the direction of $x_j$ (and vice versa). Similarly, the sensitivity due to three-factor $x_i$, $x_j$, $x_k$ interactions, (where $i \neq j \neq k$ and $1 \leq i, j, k \leq n$) is defined using a third-order partial derivative and interpreted as the change in the two-factor interaction of $x_i$ and $x_j$ as we move in the direction of $x_k$. At a given point $(x_1^*, \ \ldots, \ x_n^*)$, such two and three-factor interactions, $s_{ij}$ and $s_{ijk}$, are computed as:

$$s_{ij} = \left( \frac{\partial^2 y}{\partial x_j \partial x_i} \right)_{(x_1^*, \ \ldots, \ x_n^*)} \tag{3}$$

$$s_{ijk} = \left( \frac{\partial^3 y}{\partial x_k \partial x_j \partial x_i} \right)_{(x_1^*, \ \ldots, \ x_n^*)} \tag{4}$$

At any given point in the $n$-dimensional factor space, a total of $2^n - 1$ sensitivity measures can be calculated including $n$ sensitivity coefficients, $\binom{n}{2}$ two-factor interactions, $\binom{n}{3}$ three-factor interactions, ... , and one $n$-factor interaction. This practice of deriving point-based sensitivity measures is typically called "local sensitivity analysis," because the resulting assessment of sensitivity will, in general, only be valid in the close vicinity of the "base point" in the factor space; the exception is when the response surface is a linear function, in which case the values for the sensitivity measures remain constant across the factor space.

#### 2.2.2. LSA of Model Performance

The majority of LSA applications compute only the first-order partial derivatives of model response. However, in the context of EESMs, we are also interested in the sensitivity of one or more metrics of model *performance* to various factors, model parameters, boundary conditions, or modeling constraints etc. [e.g., *Gupta et al*., 1998, 2008, 2009].

In such cases, we have a function "optimization" problem, where the response surface (of the performance metric) will typically contain one or more stationary points at which all of the first-order partial derivatives are zero. In such cases, sensitivity can instead be characterized in terms of the matrix of *second-order partial derivatives* (the so-called *Hessian* matrix) evaluated at the stationary point where, for example $\frac{\partial^2 y}{\partial x_i^2}$ quantifies the curvature of the response surface in direction of the $i$th factor and $\frac{\partial^2 y}{\partial x_i \partial x_j}$, quantifies the interaction effects. Importantly, the *Hessian* matrix plays a critical role in the context of "gradient-based" optimization [*Bard*, 1974] and can be related to the precision of the "optimal" parameter estimates (confidence intervals) in the context of Likelihood Theory [*Edwards*, 1972].

#### 2.2.3. Numerical Approximations of the Derivatives

In a significant fraction of EESMs, their complexity makes it difficult (and expensive) to program and analytically compute the required derivatives, so it is common to estimate their values numerically via finite difference methods that approximate $\partial x_i$ by $\Delta x_i$ over some small distance. For example, the first-order sensitivity coefficient $s_i$ in equation (2) can be numerically approximated as:

$$\hat{s}_i = \left( \frac{y(x_1^*, \ \ldots, \ x_i^* + \Delta x_i, \ \ldots, \ x_n^*) \ - \ y(x_1^*, \ \ldots, \ x_i^*, \ \ldots, \ x_n^*)}{\Delta x_i} \right) \tag{5}$$

In practice, the size of $\Delta x_i$ is usually selected in an ad hoc manner. However, for significantly nonlinear responses the resulting estimate of sensitivity can depend significantly on the size of $\Delta x_i$. Further, numerical derivation of second and higher-order partial derivatives can require large numbers of model runs. In certain cases (including model calibration using performance metrics such as Mean Squared Error), the *Hessian* matrix can be approximated using computations involving only first-order partial derivatives [see *Bard*, 1974], resulting in significant computational savings.

### 2.3. Global (Population Sample-Based) Sensitivity Analysis (GSA)

Local (point-based) SA has a *unique* definition and theoretical basis, but typically provides only a limited view of model sensitivity because the results (and hence interpretation) can vary with location in the factor space. Methods of so-called "global" sensitivity analysis (GSA) attempt to provide more general results by characterizing the nature of response sensitivity over the *entire factor space*. However, this problem of

generalizing local sensitivity measures to represent "global" properties (i.e., to somehow reflect the broader characteristics of sensitivity over a domain of interest) is not trivial and, so far, no unique and definitive definition for global sensitivity exists.

For example, *Sobol'* [2001] states that global sensitivity analysis ". . .considers the model inside the box [i.e., factor space hypercube]" and ". . .should be regarded as a tool for studying the mathematical model rather than its specified solution [i.e, a base point]." *Morris* [1991] describes (global) sensitivity analysis as an experimental activity for ". . .the discovery of which inputs are important – that is, which have a substantial influence on the outputs." *Saltelli et al*. [2008] present a formal definition for (global) sensitivity analysis as "the study of how uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input." *Hamby* [1994] in his review of sensitivity analysis methods states that "many authors, when referring to the degree to which an input parameter affects the model output, use the terms 'sensitive', 'important', 'most influential', 'major contributor', 'effective', or 'correlated' interchangeably." *Hamby* [1994] also points out that of more than a dozen sensitivity analysis techniques reviewed, each provided a different sensitivity ranking; this can, of course, be traced to the lack of a unique definition for what global sensitivity actually means.

In general, methods for GSA compute their indicators of sensitivity using "values" (of something) computed at a number of different points sampled across the domain of interest (i.e., a population sample), where the sample locations are selected (in some way) to be "representative" of the entire domain; for example in Sobol Analysis, the "something" is typically a metric of model performance (a function value such as "Mean Squared Error") and a variety of strategies for generating representative samples have been proposed [*Saltelli et al*., 2008]. Therefore, GSA is to some extent rooted in the "design of experiments" (DOE), which is a broad family of statistical methods originally developed in the early 20th century [*Fisher*, 1951] for designing efficient experiments to acquire representative information when working in the context of costly, noise-prone environments. DOE has since been extended to the "design and analysis of computer experiments" (DACE), which are typically "noise-free" (also called "deterministic") in the sense that replicating a computer experiment with the same input factors results in identical model responses [*Sacks et al*., 1989]. Consistent with the GSA paradigm, DOE provides a set of tools including factorial designs and regression strategies that facilitate study of the individual and combined effects of factors on the response variable, while accounting for interaction effects [*Montgomery*, 2008]. As such, the terms "main effect" and "interaction" common in the GSA literature originate from DOE where the "main effect" of a factor is defined as the change in response due to the change in the factor *when averaged across all levels of the other factors*.

In general, the concept of GSA seems rather intuitive, although it can be interpreted differently with context and application. Broadly speaking, GSA attempts to study and quantify "how" and "to what extent" the model response *across the factor space* is influenced by a particular factor or combination of factors, either individually and/or through interactions with other factors. Since absolute sensitivity can be difficult to quantify (given that it can vary with scaling or transformation of the factor space), it is usual to focus on the *relative* sensitivity of factors with respect to each other. In section 3, different philosophies and techniques to quantify global sensitivity are discussed.

As an aside, the *sparsity-of-effects principle* indicates that most systems or processes are usually sensitive to only a subset of factors and their low-order (e.g., second-order) interactions, and that high-order interactions are typically insignificant or negligible [*Myers et al*., 2011]. While this may not be true for some models, or for all points in the factor space, the principle (particularly the assumption regarding absence of high-order interactions) has significant implications for GSA, these being increased efficiency and improved (more understandable) representation of sensitivity. Further, even if high-order effects exist, they can be difficult to interpret [*Kleijnen*, 2005], and in many cases may not have any actual physical relevance (i.e., they may be spurious artefacts of the modeling and/or system identification process).

### 2.4. Some Simple Motivational Examples

In any modeling application, the "type" of model response of interest can influence the characteristics of the associated response surface, and thereby the interpretation of a sensitivity analysis. In EESMs, methods of SA are most commonly used to investigate the effects of different factors (parameters, inputs, boundary conditions, etc.) on some model performance metric that measures goodness-of-fit between

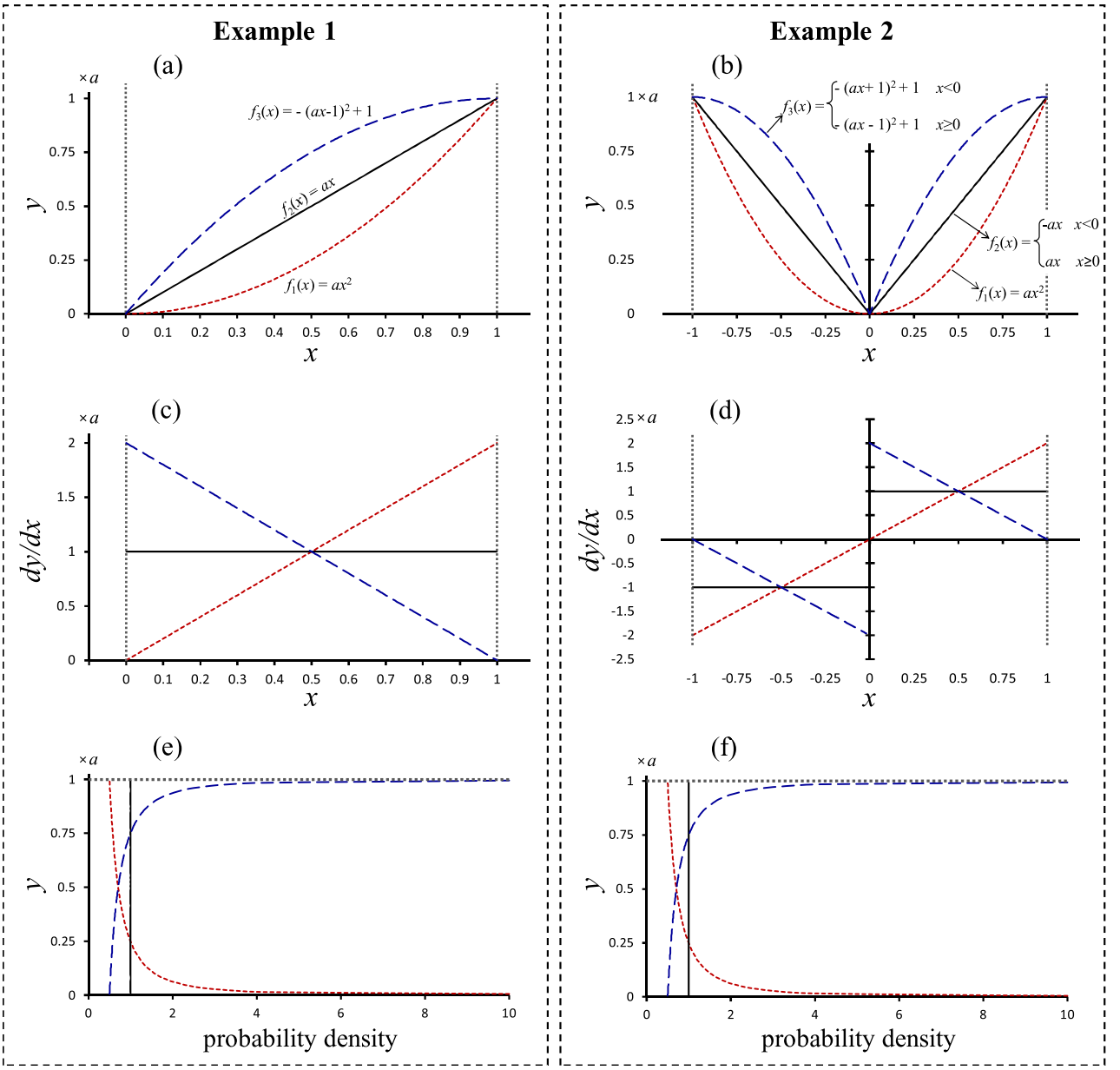


**Figure 4.** Example 1 (Monotonic Functions) and Example 2 (Uni-Modal Functions): (a) and (b) response surfaces, (c) and (d) derivative functions of the response surfaces, and (e) and (f) probability density functions of the response values.

the simulated value(s) of some system variable and corresponding observations (in some cases the metric can represent an integration over several model responses). *Duan et al.* [1992] demonstrated that this can result in (highly) nonconvex and nonsmooth response surfaces having multiple modes (i.e., "regions of attraction" in the optimization sense). In general, analysis of a complex problem of this kind will tend to focus on the region in the factor space where the model best fits the observations. In simpler cases, such as when investigating a single model response (e.g., peak flow in a rainfall-runoff model) to a small number of variables, the response surface tends to be smoother and more convex, and sensitivity over the entire factor space may be of interest.

To motivate further discussion of the conceptual basis for SA it is useful to work with some simple illustrative examples specifically designed to illustrate different issues. Accordingly, five sets of examples of *one*-factor response functions and their first-order derivatives over the factor range are depicted in Figures 4–6.

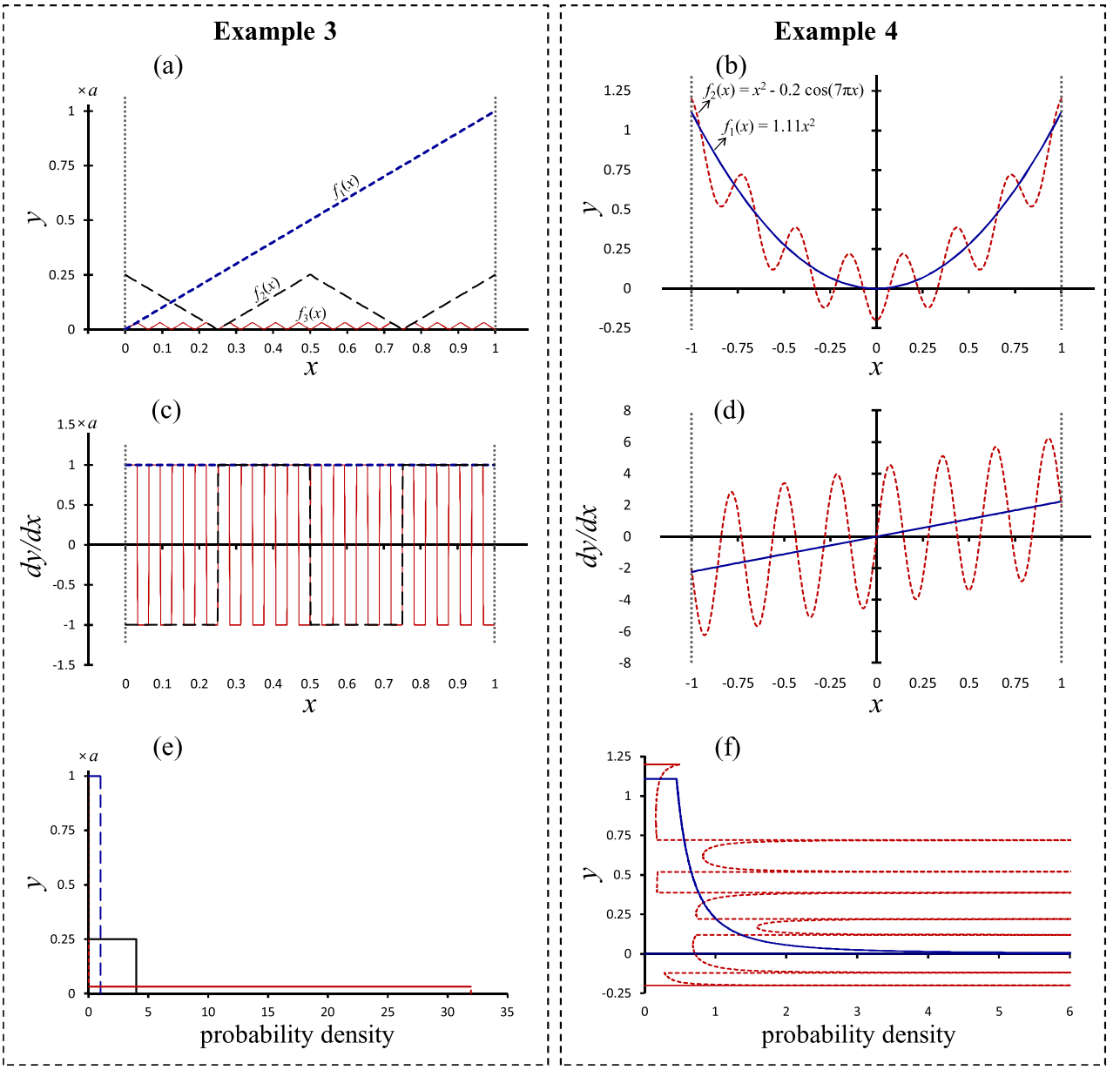


**Figure 5.** Example 3 (Functions Covering Different Ranges) and Example 4 (More Complex Functions): (a) and (b) response surfaces, (c) and (d) derivative functions of the response surfaces, and (e) and (f) probability density functions of the response values.

For each figure, we can ask the seemingly simple question "which response functions are more (or less) sensitive to variations over the range of factor x." These examples will be used to illustrate concepts of SA, and illuminate the strengths and shortcomings of existing SA approaches, thereby providing an understanding of different features and properties of response surfaces that relate to the analysis of sensitivity.

#### 2.4.1. Example 1—Monotonic Functions

Figure 4a presents three response functions that vary monotonically over the same output range. These functions have been constructed so that, although each has a different value for local sensitivity ($\partial y/\partial x$) at any given $x$ (see Figure 4c), all have the same *average* value for $\partial y/\partial x$ over the factor range. The three functions also have different probability density functions (PDFs, see Figure 4e); the PDF of a response surface represents the distribution of the response value when the factor values are assumed to be random variables with uniform distribution in their range. Note that it is not at all clear how to rank these three functions with regards to "overall sensitivity," and use of the *average* local sensitivity as an indicator of

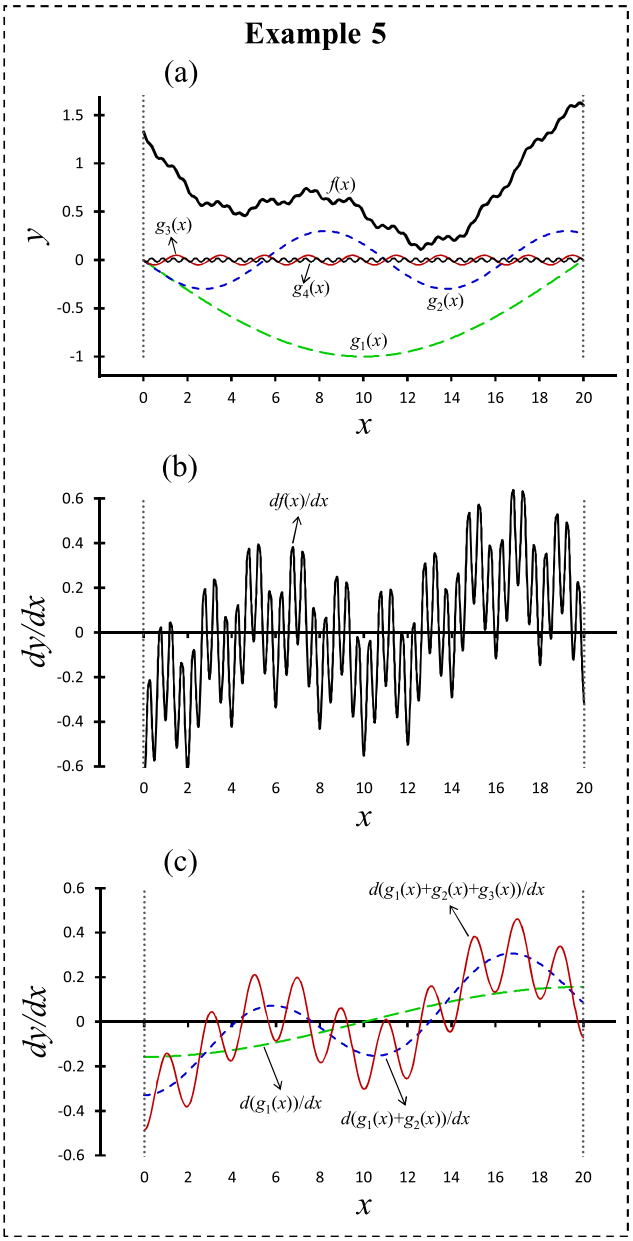


**Figure 6.** Example 5 (A Periodic Multi-Modal Function): This example demonstrates the "*scale issue*" in sensitivity analysis — (a) response surface and its constituent modes, (b) the derivative function of the response surface when $\Delta \boldsymbol{x} = dx \rightarrow 0$, and (c) example derivative functions for larger $\Delta x$ values.

"global sensitivity," would deem all the functions to be equally sensitive (average local sensitivity over the range = 1).

### 2.4.2. Example 2—Uni-Modal Functions

Figure 4b presents three uni-modal functions (each having a single unique minimum) of the kind shown in optimization textbooks (response surfaces encountered in EESM calibration will typically be more complex than these). These are idealized depictions, and the discontinuous derivative at $x = 0$ for functions $f_2$ and $f_3$ may not actually occur in modeling practice. They serve to illustrate the point that if $y$ represents how a model performance metric varies over the range of factor $x$, we would intuitively deem $f_3$ to indicate the most sensitive response since it is associated with steeper gradients in the vicinity of the function "optimum" at $x=0$. Similarly, we would identify $f_1$ as representing the least sensitive case, with $f_2$ being somewhere in between. Again, use of *average* local sensitivity as an indicator of "global sensitivity," would deem all three functions to be equally sensitive (average local sensitivity over the range = 0). Further, the PDFs of $f_1$, $f_2$, and $f_3$ in Example 2 (Figure 4f) are identical to those of functions $f_1$, $f_2$, and $f_3$ in Example 1 (Figure 4e), respectively, despite significant differences in the shape of the functions.

### 2.4.3. Example 3—Functions Covering Different Ranges

Figure 5a presents three functions that cover radically different output ranges, while having constant and identical values for *absolute* values of the local sensitivities across the factor range (except at singular points where the derivatives change sign). So all three cases have average absolute local sensitivity = 1 (Figure 5c) with significantly different response surface PDFs (Figure 5e). One possible intuitive interpretation is that $f_1$ represents the situation with highest sensitivity of the response $y$ to variations in factor $x$, because changes in $x$ control a larger range of the output. However, in some situations, periodicities (multimodalities) in the response surface may be very important in evaluating the impact of a factor. From the perspective of model calibration, the function $f_2$ (characterized by two distinct regions of attraction) might be deemed quite "sensitive" to $x$. Meanwhile $f_3$ is characterized by significant periodicity and has the exact same average absolute local sensitivity as the other two functions, but may simply represent an insensitive but nonsmooth (noisy) model response—nonsmoothness of this kind can be common in EESMs due to numerical artefacts [*Duan et al.*, 1992; *Kavetski and Clark*, 2010].

#### 2.4.4. Example 4—More Complex Functions

In a more realistic example, Figure 5b presents two functions having the same average absolute local sensitivity = 0 (Figure 5d). Despite having very different response surface PDFs (Figure 5f), they also have identical *variance* of response *y*. In this case, one might intuitively deem $f_2$ to represent the situation of higher factor sensitivity because it demonstrates both significant periodicity and variation in local sensitivity values across the *x* range. To understand this, ask yourself the question "which function represents the situation where the value of y changes more rapidly as x is varied over its range?" If one is concerned with the local stability of the response to small changes in factor *x* then clearly function $f_2$ represents the situation of higher sensitivity, but if one is concerned mainly with the broad overall change in *y* over the total range of *x*, then both functions can be deemed to be fairly similar in that respect.

#### 2.4.5. Example 5—A Periodic Multimodal Function

To conclude our motivational illustration of the difficulty of defining what is meant by sensitivity, our final example is based on the fact that any periodic (multimodal) response surface can be decomposed (e.g., via Fourier series analysis) into a set of simpler periodic functions having different amplitudes and frequencies. Figure 6a shows a periodic function $f(x)$, and its decomposition into four constituent modes:

$$f(x)=1.33-\sin\left(\frac{2\pi x}{40}\right)-0.3\sin\left(\frac{2\pi x}{11}\right)-0.05\sin\left(\frac{2\pi x}{2}\right)-0.02\sin\left(\frac{2\pi x}{0.5}\right) \tag{6}$$

where the sinusoidal terms correspond respectively to functions $g_1(x)$, $g_2(x)$, $g_3(x)$, and $g_4(x)$. Figure 6b shows the analytical derivative of $f(x)$ with respect to factor *x*. Clearly, the component function $g_4(x)$ dominates the contribution to local sensitivity (slope) across the factor range, and could lead an analyst to infer a very large sensitivity. However, if $g_4(x)$ primarily represents noise due to data errors or numerical artefacts, such an interpretation can be highly misleading.

In practice, partial derivatives (and associated sensitivity measures) are typically calculated by finite difference procedures on the basis of some (often arbitrarily chosen) step size $\Delta x$, as shown in equation (5). In such situations, the selection of larger $\Delta x$ values will cause the analysis to be less sensitive to high-frequency roughness in the response surface—a fact exploited by the Shuffled Complex Evolution optimization method of *Duan et al*. [1992]. Figure 6c illustrates this for $\Delta x$ equals 0.5 and 2 (corresponding to the periods of the $g_4(x)$ and $g_3(x)$ waves) and for which the derivative functions are then equivalent to $d(g_1(x)+g_2(x)+g_3(x))/dx$ and $d(g_1(x)+g_2(x))/dx$, respectively (when $\Delta x=11$, the derivate function is close to $d(g_1(x))/dx$ shown in Figure 6c). Clearly, interpretation of the sensitivity of $f(x)$ with respect to variations in *x* can be very different for different $\Delta x$ values. While this issue has not attracted much attention in the SA literature, it is clearly of relevance and concern since the response surfaces of typical simulation models in the Earth and Environmental Sciences have similar characteristics [*Duan et al*., 1992]. In our subsequent discussion, we will refer to this as the "scale issue."

## 3. Historical Evolution of GSA Methods

The earliest approaches to SA were based largely in the use of partial derivatives (see section 2.2) and the concept of a Taylor series expansion around a base point [*Hamby*, 1994; *Rabitz et al*., 1983], while derivative-based LSA methods also have roots in optimization theory. Over the past several decades, attempts to represent the more "global" nature (over the factor domain as opposed to at a specific point) of the sensitivity of model response have been made. Early attempts were based in concepts developed for the statistical design of experiments, including the one-factor-at-a-time (OAT) method, factorial design, and regression and correlation analysis. These classic methods, referred to as model-based methods in the following, *assume* a particular model form (typically linear or polynomial) for the underlying response surface. Model-free GSA methods, however, are more general and have evolved in the context of computer experiments.

### 3.1. Model-based Methods

#### 3.1.1. One-Factor-At-a-Time Method (OAT)

The OAT method (probably the simplest of classic techniques) computes a finite difference approximation of the local slope of the response surface around a base point in the factor space. In practice, because the size $\Delta x$ of the factor change in OAT is typically some fraction (e.g., 1–10%) of the factor range, the method actually detects larger-scale trends (lower frequency variations) in the response surface. Further, OAT does

not detect and measure factor interactions. Nonetheless, the approach is computationally efficient, requiring only $n+1$ model runs for an $n$-dimensional factor space. *Saltelli and Annoni* [2010] note that use of OAT is very common in the modeling community, and raise fundamental concerns regarding the validity of sensitivity results based on it.

#### 3.1.2. Factorial Design

In "full factorial" designs, the factor space is discretized into a certain number of levels (i.e., grid points) along the direction of each factor and the model response is computed for each combination of factor levels. These points are then used to estimate the "main effects" of each factor (providing a global measure of first-order sensitivity) and also the "interaction effects" between various combinations of factors (providing global measures of second and higher-order sensitivities). The degree to which this approach properly represents global sensitivity depends on a number of things, including the selected grid spacing (number of levels for each factor) and the degree of nonlinearity in the underlying response. Because an $m$-level full factorial design in an $n$-dimensional factor space requires $m^n$ model runs, the approach is subject to the *curse of dimensionality* and computational cost can become prohibitive as problem dimension increases. To mitigate this latter issue, *fractional* factorial designs that rely on the sparsity-of-effects principle have been proposed; these use carefully selected subsamples of the full factorial design to estimate only the main and low-order interaction effects (see *Kleijnen* [2005] for a comprehensive review).

#### 3.1.3. Response Surface Approximation via Regression

A variety of regression techniques have been used to approximate response surfaces for the purpose of SA, generally using either linear (with or without second-order interaction terms) or quadratic polynomials; such approaches are sometimes called "response surface methods." Once a regression model is fit to a set of points sampled in the factor space (note the connection to factorial design), the coefficients of the regression equation can be interpreted as indices of factor sensitivity—the coefficients of linear and second-order interaction terms correspond to the main and interaction effects respectively, while second-order terms can be used to detect nonlinearities in the response. To sample the factor space, both deterministic and random sampling techniques (e.g., factorial design and Latin hypercube) have been used.

A major drawback of the regression-based approach is its heavy reliance on the prior assumption regarding model form (i.e., regression equation), and if the regression equation does not fit the underlying response surface well, the sensitivity estimates can be seriously incorrect. Rank transformation is a common way to partially address the nonlinearity of response surfaces [*Iman and Conover*, 1979]; however, such an approach may only be useful in case of monotonic response surfaces and fail in the presence of multimodality. For a comprehensive review of regression techniques for GSA see *Kleijnen* [1995].

### 3.2. Model-Free Methods

Historically, the methods described above were originally developed for analysis of physical experiments. In the context of computer-based modeling, it became affordable to conduct simulation experiments, and led to development of methods requiring large numbers of samples. Such methods were also motivated by the fact that the degree of convexity of the underlying response surface of a complex model cannot generally be known a priori, raising the need for "model-free" methods that make minimal assumptions regarding the functional form of the response surface. The rest of this section is dedicated to a brief summary of such methods.

#### 3.2.1. Regional Sensitivity Analysis (RSA)

A heuristic, but commonly used approach is to partition the marginal distribution of samples obtained for each factor (typically a bounded uniform distribution) into two (or more) distributions based on empirically selected threshold values for model response. The idea is that if the factor does not have a significant impact on model response throughout the factor space, the two distributions should not be statistically distinguishable. Formal statistical tests (such as the *Smirnov* test) can be used to quantify the extent to which two distributions are significantly different, and provide a level of significance used for factor ranking. In the literature, this procedure is variously referred to as "regional sensitivity analysis" [*Spear et al*., 1994], "sensitivity tests involving segmented input distributions" [*Hamby*, 1994], or "Monte-Carlo filtering" [*Saltelli et al*., 2008]. For a review of the approach and different methods for statistical testing see *Hamby* [1994].

#### 3.2.2. Variance-Based Methods

Perhaps the most sophisticated approach developed to-date, for defining and quantifying "global" sensitivity, is "variance-based" SA, based on the idea that the *variability* of model response in each factor direction can be used as a direct indicator of factor sensitivity. The contribution of each factor (main or first-order effect), and each possible two or higher-order interaction (interaction effect), to the total variance of the model response is quantified, and the ratio of each contribution to the total variance is interpreted as the measure of sensitivity.

##### 3.2.2.1. Theoretical Basis

Given the model response function presented in equation (1), the variance $V(y)$ of the model response can be theoretically decomposed to $2^n-1$ components as follows:

$$V(y)=\sum_{i=1}^{n} V_i+\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} V_{ij}+\sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\sum_{k=j+1}^{n} V_{ijk}+\ldots+V_{1\ldots n} \tag{7}$$

where $V_i$ is the contribution of the $i^{\text{th}}$ factor to the total variance excluding its interactions with other factors, $V_{ij}$ is the contribution of the two-factor interaction of the $i^{\text{th}}$ and $j^{\text{th}}$ factors to the total variance, $V_{ijk}$ is the contribution of the three-factor interaction of the $i^{\text{th}}$, $j^{\text{th}}$, and $k^{\text{th}}$ factors to the total variance, and so on.

These components can be computed as $V_i=V(E(y|x_i))$, $V_{ij}=V\left(E\left(y|x_i,x_j\right)\right)-V_i-V_j$, and $V_{ijk}=V\left(E\left(y|x_i,x_j,x_k\right)\right)-V_i-V_j-V_k-V_{ij}-V_{ik}-V_{jk}$, where $E(y|x_i)$ indicates the expected value of $y$ given $x_i$, and so on. Accordingly, the associated first, second, and third-order sensitivity indices are defined as $S_i=\frac{V_i}{V(y)}$, $S_{ij}=\frac{V_{ij}}{V(y)}$, and $S_{ijk}=\frac{V_{ijk}}{V(y)}$.

Of particular interest in variance-based SA is the "total-order effect" sensitivity index, $S_{Ti}$, which sums over the first-order effect of the $i$th factor and its interactions of any order with any other factors. To circumvent having to compute all the related terms, the total-effect sensitivity index can be efficiently calculated as follows:

$$S_{Ti}=1-\frac{V(E(y|x_1,x_2,\ldots,x_{i-1},x_{i+1}\ldots,x_n))}{V(y)} \tag{8}$$

where the numerator consists of all the terms of any order that *do not include* the $i^{\text{th}}$ factor. A thorough description of the variance-based approach is available in *Saltelli et al.* [2008].

A very appealing feature of variance-based methods is their ability to explicitly differentiate between interaction and main effects. Whereas EESMs may contain factors (e.g., parameters) that are significantly correlated, GSA approaches commonly assume the prior distributions of factors to be uncorrelated (i.e., orthogonal). The variance-based methods facilitate a posterior assessment of the extent and strength of correlations (i.e., interaction effects) between different factors. The prior assumption of an orthogonal factor space is consistent with the common practice of Bayesian uncertainty analysis for EESMs. Combined with this assumption, equation (7) indicates that the total-order effect of a factor is always equal to or greater than the main effect of that factor – which runs contrary to suggestions in the literature that correlation effects can cause the total-order effects in a variance-based analysis to become smaller than main effects [e.g., *Pappenberger et al.*, 2008].

For completeness, we note that multiple-criteria extensions of the variance-based GSA approach have also been reported [see *Bastidas et al.*, 2006; *Gupta et al.*, 1999; *Rosolem et al.*, 2012]. However, we will not discuss these approaches further here.

##### 3.2.2.2. Computational Implementation

For simple, analytically tractable functions, variance-based sensitivity indices can be calculated analytically. The Fourier Amplitude Sensitivity Test (FAST), developed by *Cukier et al.* [1973] and promoted by *Saltelli and Bolado* [1998], provides an efficient way to numerically compute variance-based sensitivity indices. However, FAST can provide only first-order sensitivity indices, and a later development called the "extended FAST" (EFAST) allows computation of total effects [*Saltelli et al.*, 1999].

Perhaps the most popular implementation of variance-based SA is the so-called "Sobol method," named after Ilya M. Sobol' [*Sobol'*, 1990], which provides numerical estimates of first, higher-order, and total-effect sensitivity indices in a relatively efficient manner based on Monte-Carlo sampling [*Homma and Saltelli*, 1996]. For a detailed description, see *Saltelli et al.* [2008].

#### 3.2.2.3. Entropy-Based Methods

Given that the second-order moment (i.e., variance) of a response surface may not be sufficient to adequately characterize the distributional properties (and uncertainty) of the model response, there have also been efforts to utilize "entropy-based" criteria that quantify the extent to which a response distribution differs from a uniform distribution. Like variance-based methods, this approach is based on a criterion that characterizes the distributional properties of a response surface, and so can be thought of as an extension (or refinement) of the variance-based approach. The first report on this approach seems to be *Krzykacz-Hausmann* [2001]. A review is provided by *Liu et al*. [2006].

### 3.2.3. Globally Aggregated Measures of Local Sensitivities

Methods for globalized evaluation of local sensitivities (i.e., first-order partial derivatives – see equation (2)) have also been developed. Such methods evaluate the local sensitivity coefficients for each factor at multiple points across the factor space, and analyze the distributional properties of these values to assess the global sensitivities of model response to individual factors. The seminal paper by *Morris* [1991] proposed evaluation of the following terms with respect to factor $x_i$ ($i = 1, \ldots, n$):

$$\mu_i = E\left(\frac{\partial y}{\partial x_i}\right) \tag{9}$$

$$\sigma_i^2 = V\left(\frac{\partial y}{\partial x_i}\right) \tag{10}$$

and proposed an experimental design for numerically evaluating the distribution of local sensitivities (which he referred to as "elementary effects") across the factor space. For any given problem, larger values of $\mu_i$ and $\sigma_i$ generally indicate higher sensitivities to the associated factor. *Campolongo et al*. [2007] pointed out the limitations of this approach when the response surface is nonmonotonic, and proposed the use of $\mu_i^*$ instead of (or along with) $\mu_i$ as defined below:

$$\mu_i^* = E\left(\left|\frac{\partial y}{\partial x_i}\right|\right) \tag{11}$$

*Campolongo et al*. [2007] further tried to empirically demonstrate a direct relationship between $\mu_i^*$ and $S_{Ti}$, the total-order effect sensitivity index defined in equation (8). *Sobol and Kucherenko* [2009] showed that the interpretation of global sensitivities can be improved by using the squared partial derivatives rather than their absolute values, as defined by:

$$v_i = E\left(\left(\frac{\partial y}{\partial x_i}\right)^2\right) \tag{12}$$

and established a theoretical relationship between $v_i$ and $S_{Ti}$, which identifies an upper bound for $S_{Ti}$ as a function of $v_i$. *Rakovec et al*. [2014] recently proposed a similar criterion (in the DELSA method) which is effectively the same as equation (12).

So in these approaches, global sensitivity analysis is based on the numerical evaluation and interpretation of these criteria, which are "variations" of the Morris method, and are sometimes considered as efficient proxies for variance-based sensitivity analysis [*Campolongo et al*., 2007; *Sobol and Kucherenko*, 2009]. However, *Sobol and Kucherenko* [2009] demonstrate that ranking of influential variables based on these criteria may result in false conclusions (provided one assumes that the variance-based total-effect sensitivity index provides a *correct* assessment). They also demonstrate that when the "characteristic length of function variation" is smaller than the discretization step size used for numerical evaluation of local sensitivities, the resulting assessment can be inaccurate – a reference to the "scale issue" raised in section 2.4.5.

## 3.3. Methods Quantifying "Global" Interaction Effects

Originating in the design of experiments and regression analysis, the concept of factor interaction indicates a situation where the simultaneous effects of two or more factors (independent variables) on a response (dependent variable) are *nonadditive*. When defined at a single point on the response surface, the interaction effects between any subset of factors can be directly quantified and interpreted through second and higher-order partial derivatives (see section 2.2). However, generalizing these local (point-based) measures

of interaction to a global measure that characterizes the entire factor space is nontrivial and, to our knowledge, no attempts to directly extend the local interaction measures to global scale have been reported (whereas extensions to first-order sensitivities have been developed, as discussed in section 3.2.3). This lack may be partly due to the large computational demands required to numerically calculate higher-order partial derivatives.

The classic techniques of factorial design (see section 3.1.2) and regression analysis (see section 3.1.3) are relatively simple ways to obtain approximate information about the "global interaction effects" (hereafter referred to as "interaction effects") across the factor space. For (almost) linear response surfaces, these methods provide (approximate) assessments of second and higher-order partial derivatives. Variance-based methods, which are conceptually different at a fundamental level, are perhaps the most advanced techniques to quantify interaction effects (in some sense), by calculating the portion of the total variance of the response surface that is due to the interaction effect of interest (see section 3.2.2). Nonetheless, the calculation of interaction effects can be computationally challenging.

Beyond the computational issue, there exists the more significant challenge of *interpreting* the resulting interaction effects. This problem can be nontrivial, even for simple two-factor response surfaces, $y=f(x_1,x_2)$. To illustrate the difficulty of interpreting interaction effects, we analyze six example two-factor response surfaces in section 4.2. These examples provide insight into the challenges involved in interpreting the results of different methods for estimating interaction effects in real multidimensional EESMs.

## 4. Shortcomings of Existing Approaches

### 4.1. Response Surfaces With No Interactions

The *Sobol* and *Morris* methods for evaluating sensitivity have received much attention from the community. Using the simple examples presented in section 2.4, we demonstrate that these methods suffer from serious shortcomings and do not always provide results that are consistent with an intuitive understanding of sensitivity; note that for each of these examples, the variance $V(y)$ of the response surface can be computed analytically, as can the Morris indices $E(|\frac{\partial y}{\partial x}|)$ and $V\left(\frac{\partial y}{\partial x}\right)$, and the variation $E\left(\left(\frac{\partial y}{\partial x}\right)^2\right)$ presented above. Table 1 reports results for the response surfaces of Examples 1–5 (Figures 4–6).

#### 4.1.1. Example 1

The three monotonically increasing functions $f_1(x)$, $f_2(x)$ and $f_3(x)$ shown in Figure 4a have different shapes. However, all of the measures suggest that $f_1(x)$ and $f_3(x)$ are equally sensitive to factor $x$, while the linear function $f_2(x)$ is equally sensitive according to measure $E(|\frac{\partial y}{\partial x}|)$, and slightly less sensitive according to measures $V(y)$ and $E\left(\left(\frac{\partial y}{\partial x}\right)^2\right)$.

**Table 1.** Performance of the Conventional Sensitivity Measures on Examples 1–5

| | | Sensitivity Measures | | | |
|---|---|---|---|---|---|
| Hypothetical Response Surfaces | | $V(y)$ | $E(\lvert\frac{\partial y}{\partial x}\rvert)$ | $E\left(\left(\frac{\partial y}{\partial x}\right)^2\right)$ | $V\left(\frac{\partial y}{\partial x}\right)$ |
| Example 1* Example 2** | $f_1(x)$ | $\frac{4a^2}{45}$ | $a$ | $\frac{a^2}{3}$ | $\frac{a^2}{3}$* and $\frac{4a^2}{3}$** |
| | $f_2(x)$ | $\frac{a^2}{12}$ | $a$ | $a^2$ | 0* and $a^2$** |
| | $f_3(x)$ | $\frac{4a^2}{45}$ | $a$ | $\frac{4a^2}{3}$ | $\frac{a^2}{3}$* and $\frac{4a^2}{3}$** |
| Example 3 | $f_1(x)$ | $\frac{a^2}{12}$ | $a$ | $a^2$ | 0 |
| | $f_2(x)$ | $\frac{a^2}{192}$ | $a$ | $a^2$ | $a^2$ |
| | $f_3(x)$ | $\frac{a^2}{12288}$ | $a$ | $a^2$ | $a^2$ |
| Example 4 | $f_1(x)$ | 0.11 | 1.11 | 1.64 | 1.64 |
| | $f_2(x)$ | 0.11 | 3.01 | 11.80 | 11.80 |
| Example 5 | $f(x)$ | 0.141 | 0.22 | 0.08 | 0.08 |
| | $f(x)-g_4(x)$ | 0.141 | 0.17[a] | 0.04[a] | 0.04[a] |
| | $f(x)-g_3(x)-g_4(x)$ | 0.139 | 0.15[b] | 0.03[b] | 0.03[b] |

[a]Equivalent to the numerical analysis of $f(x)$ with discretization of $\Delta x=0.5$.
[b]Equivalent to the numerical analysis of $f(x)$ with discretization of $\Delta x=2$.

**Table 2.** Performance of Direct and Interaction Effects Sensitivity Measures on Examples 6–11

| | Mean Local Measures | | | Mean Squared Local Measures | | | Variance-Based Measures[a] | | |
|---|---|---|---|---|---|---|---|---|---|
| Hypothetical Response Surfaces | $E\left(\frac{\partial y}{\partial x_1}\right)$ | $E\left(\frac{\partial y}{\partial x_2}\right)$ | $E\left(\frac{\partial^2 y}{\partial x_1 \partial x_2}\right)$ | $E\left(\left(\frac{\partial y}{\partial x_1}\right)^2\right)$ | $E\left(\left(\frac{\partial y}{\partial x_2}\right)^2\right)$ | $E\left(\left(\frac{\partial^2 y}{\partial x_1 \partial x_2}\right)^2\right)$ | $V_1$ | $V_2$ | $V_{1,2}$ |
| Example 6: $y = ax_1 + bx_2 + cx_1x_2$ $x_1$ & $x_2 \in [0, 1]$ | $a+c/2$ | $b+c/2$ | $c$ | $a^2+ac+c^2/3$ | $b^2+bc+c^2/3$ | $c^2$ | $(a+c/2)^2/12$ | $(b+c/2)^2/12$ | $c^2/144$ |
| Example 7: $y = cx_1x_2$ $x_1$ & $x_2 \in [-1, 1]$ | 0 | 0 | $c$ | $c^2/3$ | $c^2/3$ | $c^2$ | 0 | 0 | $c^2/9$ |
| Example 8: $y=ax_1^2+bx_2^2+cx_1x_2$ $x_1$ & $x_2 \in [-1, 1]$ | 0 | 0 | $c$ | $4a^2/3+c^2$ | $4b^2/3+c^2$ | $c^2$ | $4a^2/45$ | $4b^2/45$ | $c^2/9$ |
| Example 9: $y=ax_1^2+bx_2^2+cx_1^2x_2$ $x_1$ & $x_2 \in [-1, 1]$ | 0 | $c/3$ | 0 | $4a^2/3+4c^2/9$ | $4b^2/3+c^2/5$ | $4c^2/3$ | $4a^2/45$ | $4b^2/45+c^2/27$ | $4c^2/135$ |
| Example 10[b]: Uncorrelated normal distribution with $\sigma_{x_1}^2=0.075$ and $\sigma_{x_2}^2=0.05$ | 0 | 0 | 0 | 0.32 | 0.48 | 3.17 | 0.010 | 0.014 | 0.014 |
| Example 11[b]: Correlated normal distribution with $\rho=0.33$ $\sigma_{x_1}^2=0.075$ and $\sigma_{x_2}^2=0.05$ | 0 | 0 | $10^{-6}$ | 0.33 | 0.50 | 4.56 | 0.009 | 0.013 | 0.015 |

[a]The total variance of y, V(y), equals $V_1+V_2+V_{1,2}$. Total-order effects of $x_1$ and $x_2$ are $V_1+V_{1,2}$ and $V_2+V_{1,2}$, respectively.
[b]The response surface is scaled to the *y* range [0–1].

#### 4.1.2. Example 2

Results for the functions $f_1(x)$, $f_2(x)$ and $f_3(x)$ having single global minima shown in Figure 4b are identical to those reported for the monotonic functions in Figure 4a. From a model calibration perspective, we would intuitively consider $f_1(x)$ to represent the least sensitive, and $f_3(x)$ to represent the most sensitive, of the three cases. Because the PDFs of $f_1(x)$, $f_2(x)$, and $f_3(x)$ in Example 2 are respectively the same as the PDFs of $f_1(x)$, $f_2(x)$, and $f_3(x)$ in Example 1, no PDF-based sensitivity approach, including an entropy-based approach, can differentiate between them.

#### 4.1.3. Example 3

For the response surfaces illustrated in Figure 5a, the measure $V(y)$ assesses function $f_1(x)$ as being 16 and 1026 times more sensitive to factor $x$ than functions $f_2(x)$ and $f_3(x)$, respectively. However, measures $E(|\frac{\partial y}{\partial x}|)$ and $E\left(\left(\frac{\partial y}{\partial x}\right)^2\right)$ suggest that all functions are equally sensitive, while measure $V\left(\frac{\partial y}{\partial x}\right)$ does not distinguish between $f_2(x)$ and $f_3(x)$. If these functions were to represent problems of function minimization, the results provided by $V(y)$ would be more consistent with intuition.

#### 4.1.4. Example 4

Although the two response surfaces shown in Figure 5b have different length scales of variation, they have identical variances V(y), which runs counter to our intuitive notion of sensitivity. This happens because the variance approach measures overall variability of the response but is not sensitive to the *structure* of the surface—i.e., how the values of the response surface are organized in the factor space. Consequently, variance-based methods are unable to take into account important structural information such as multimodality. In contrast, the Morris type methods indicate that $f_2(x)$ is significantly more sensitive to factor $x$ than is $f_1(x)$.

#### 4.1.5. Example 5

The function shown in Figure 6a is constructed to illustrate the impact of roughness (high-frequency low-amplitude noise) on the performance of sensitivity metrics. The variance $V(y)$ of $f(x)$, is almost identical to that of $f(x)-\ g_4(x)$ and $f(x)-\ g_3(x)-\ g_4(x)$, which are basically $f(x)$ with the high-frequency periodicities removed, showing that variance-based approach to be robust in the presence of roughness (high frequency and low amplitude noise). Conversely, the Morris type methods indicate significantly different sensitivities for the three cases. However, the Morris methods provide identical results for $f(x)$ and $f(x)-\ g_4(x)$ when the derivatives are computed numerically using $\Delta x = 0.5$, and for $f(x)$ and $f(x)-\ g_3(x)-\ g_4(x)$ when the derivatives are computed numerically using $\Delta x = 2$, illustrating lack of robustness to choice of discretization step size, and hence to the scale issue raised in section 2.4.5.

### 4.2. Response Surfaces With Interactions

As for the global sensitivity indices obtained through benchmark GSA methods, the interpretation of global interaction effects can be challenging and counter-intuitive. This is true even for simple two-factor response

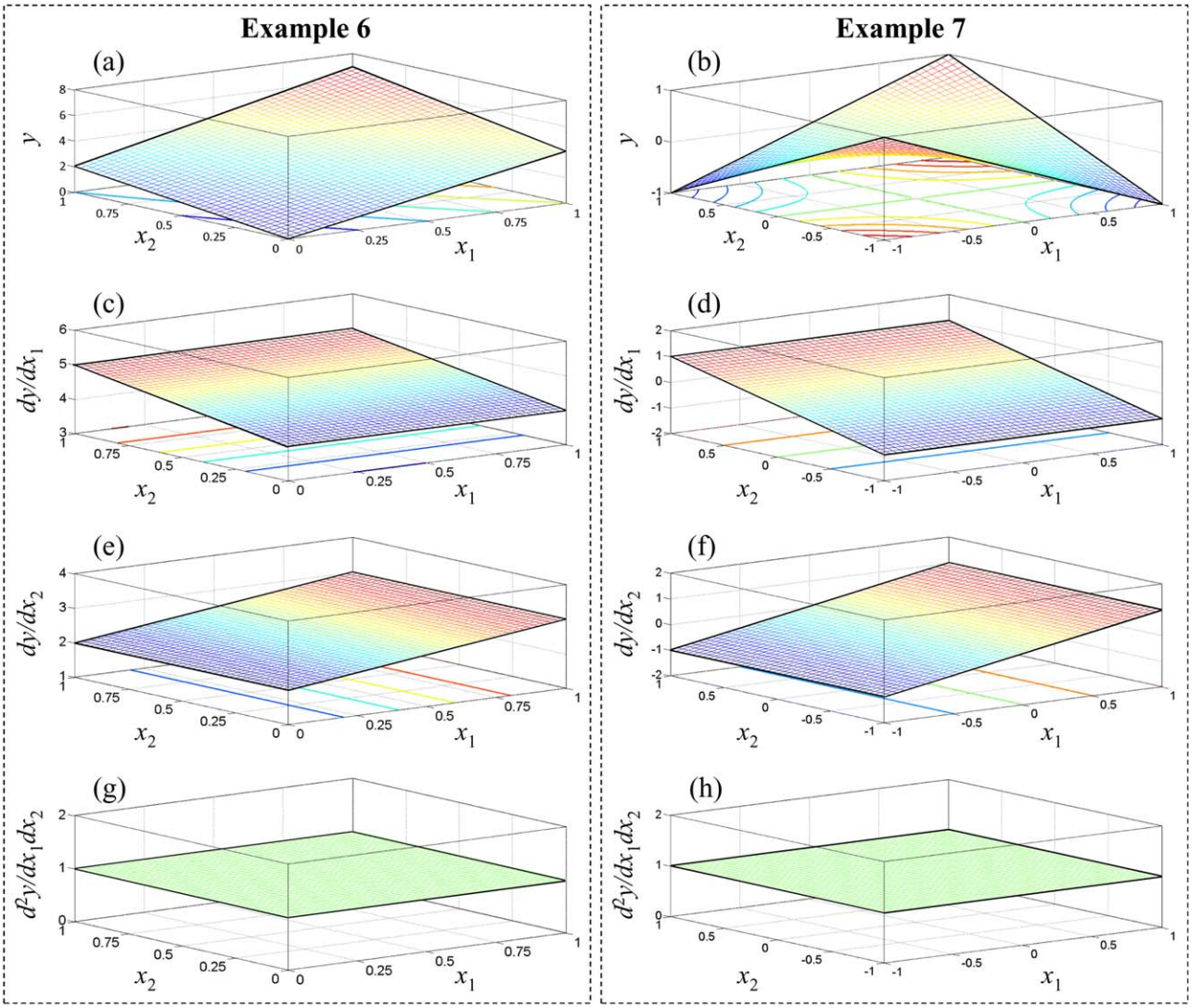


**Figure 7.** Examples 6 and 7 – Linear and constant response surfaces augmented with constant interaction term: (a) and (b) response surfaces, (c) and (d) surfaces representing partial derivatives with respect to $x_1$, (e) and (f) surfaces representing partial derivatives with respect to $x_2$, (g) and (h) surfaces representing second-order derivatives. Example 6 follows equation $y = ax_1 + bx_2 + cx_1x_2$ where a = 4, b = 2, and c = 1. Example 7 follows equation $y = cx_1x_2$ where c = 1.

surfaces, $y=f(x_1, x_2)$, as demonstrated below using six simple examples (Table 2 and Figures 7–10) that help illustrate how different methods quantify the interactions between factors (as well as the "direct effect" of each factor). As before, we are interested in whether the results of different methods are consistent with our intuitive understanding of interaction.

Table 2 presents the results of three methods used to characterize the direct and interaction effects across the factor space. The first method averages $\frac{\partial y}{\partial x_1}$, $\frac{\partial y}{\partial x_2}$, and $\frac{\partial^2 y}{\partial x_1 \partial x_2}$ over the entire factor space (i.e., expectations of the local measures). The second method is similar but averages the squared values (i.e., expectations of the squared local measures). The third method is the variance-based approach explained in section 3.2.2.

#### 4.2.1. Example 6

Figure 7a shows a linear function augmented by *a constant interaction* term $cx_1x_2$. In this case, the second-order derivative is constant across the factor space (i.e., $\frac{\partial^2 y}{\partial x_1 \partial x_2} = c$), see Figure 7g, and provides an intuitively justified measure of the interaction effect. Note that the significance of an interaction effect is evident when benchmarked in a relative sense against the direct effects. The two other measures provide a different assessment. To highlight the difference, consider the case where a=0, when the ratio of the interaction

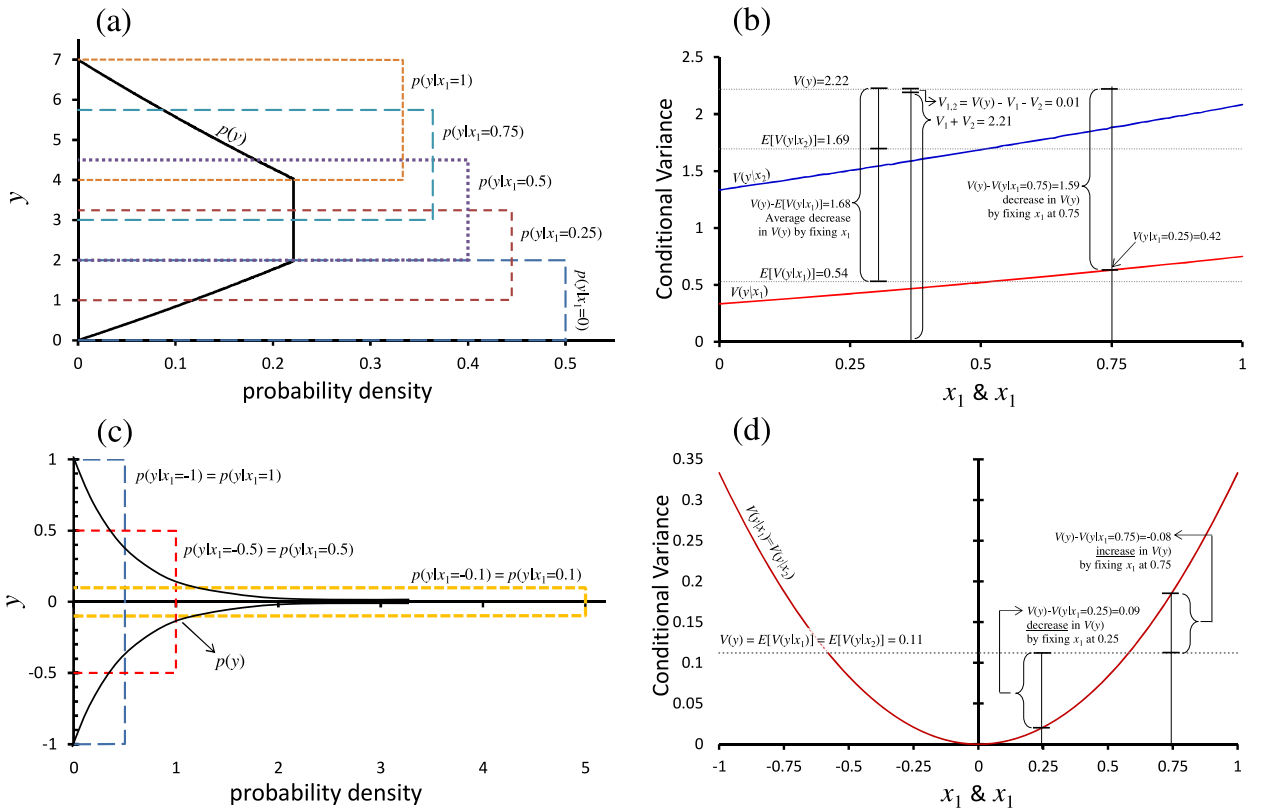


**Figure 8.** Example illustrations of how the variance-based method works on Examples 6 and 7. Example 6: (a) unconditional and conditional probability density functions of *y*, (b) unconditional and conditional variance of *y*. Example 7: (c) unconditional and conditional probability density functions of *y*, (d) unconditional and conditional variance of *y*.

effect to the direct effect of $x_1$ is 2 for the first measure and 3 for the second and third measures. For $a \neq 0$, the ratios for the three measures are different.

#### 4.2.2. Example 7

Figure 7b shows a function having only a constant interaction term, with $x_1$ and $x_2$ varying on the range [−1 to +1]. The direct effects of $x_1$ and $x_2$ assessed by the first and third methods are zero, and as such, all variation in the response surface is attributed to the interaction effect. This is despite the fact that the local sensitivity with respect to $x_1$ or $x_2$ is nonzero at any given point. In this example, if $x_1$ and $x_2$ were varied on the range [0 to +1], then all the three methods would provide nonzero assessments for the direct and interaction effects.

Figure 8 further demonstrates how the variance-based method works when applied to Examples 6 and 7. For Example 6, Figure 8a shows the PDF of *y*, labeled $p(y)$, when $x_1$ and $x_2$ are assumed to be uniformly distributed across the factor space, as well as the conditional PDFs of *y*, labeled $p(y|x_1)$, when $x_1$ is fixed at certain values in its range. The change in conditional PDF $p(y|x_1)$ from that obtained by fixing $x_1$ at its true value can be thought of as characterizing the sensitivity of *y* with respect to $x_1$.

Let $V(y)$ represent the variance of the response surface computed from $p(y)$. Figure 8b shows the conditional variances $V(y|x_1)$ and $V(y|x_2)$ when $x_1$ and $x_2$ are fixed at different values. When $x_1$ is fixed at 0.75, the reduction of variance due to this fixing is given by $V(y) - V(y|x_1=0.75)$. Since the true value of $x_1$ is unknown, the variance-based approach computes instead the *average* reduction in variance for any possible value of $x_1$, given by $V(y) - E[V(y|x_1)]$, and uses this to represent the first-order sensitivity (main effect) of *y* to factor $x_1$.

Now according to the law of total variance $V(y) = E[V(y|x_1)] + V[E(y|x_1)]$, and so an alternative way to compute the main effect of factor $x_1$ is given by $V_1 = V[E(y|x_1)]$ (as used in section 3.2.2). Similarly the main effect of factor $x_2$ is given by $V_2 = E[V(y|x_2)]$. Knowing these, the portion of the total variance attributable to the interaction effect, called $V_{1,2}$, is obtained as $V(y) - V_1 - V_2$.

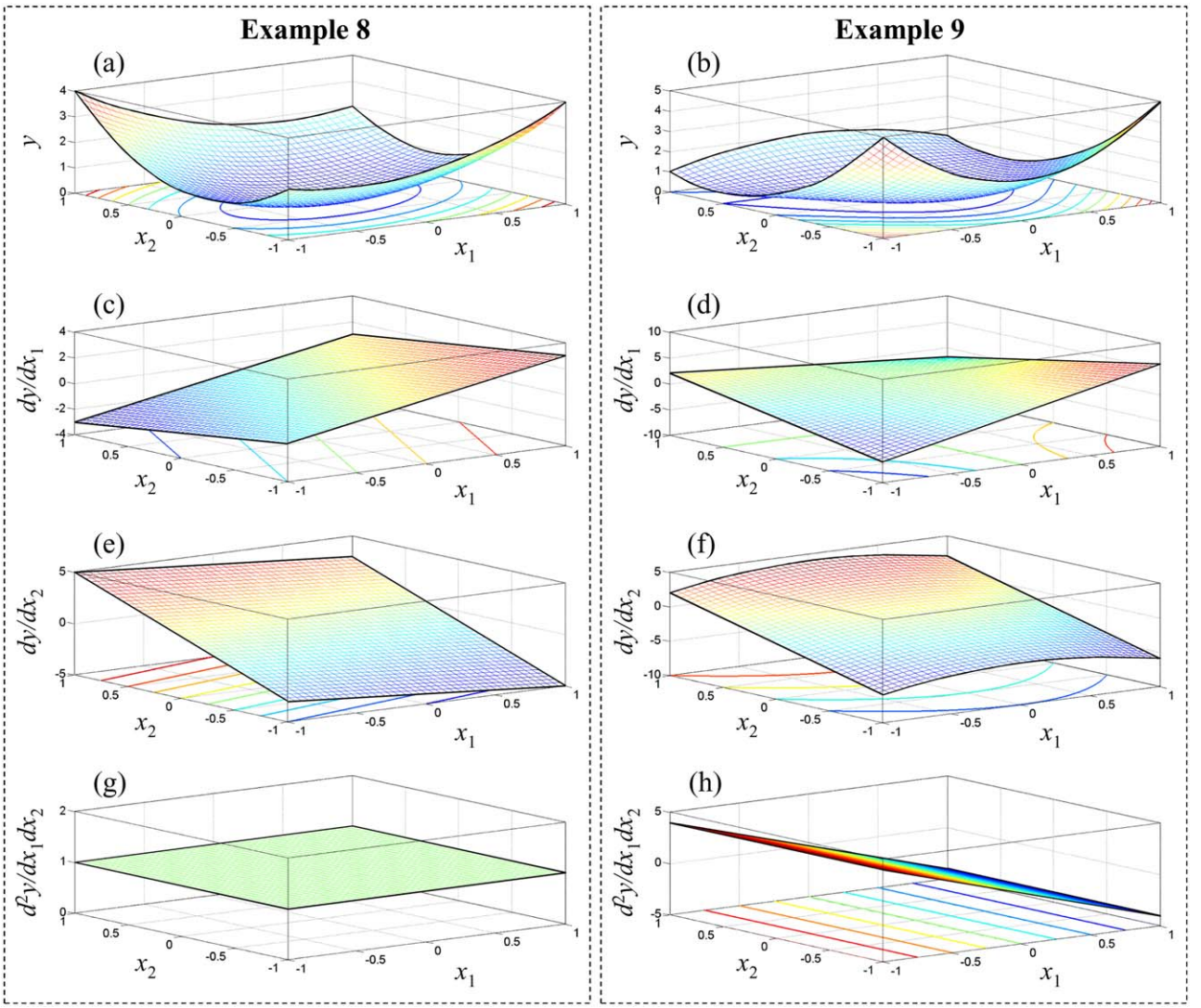


**Figure 9.** Examples 8 and 9 – Quadratic response surfaces augmented with constant and variable interaction terms: (a) and (b) response surfaces, (c) and (d) surfaces representing partial derivatives with respect to $x_1$, (e) and (f) surfaces representing partial derivatives with respect to $x2$, (g) and (h) surfaces representing second-order derivatives. Example 8 follows equation $y = ax_1^2 + bx_2^2 + cx_1x_2$ where a = 1, b = 2, and c = −1. Example 9 follows equation $y = ax_1^2 + bx_2^2 + cx_1^2x_2$ where a = 1, b = 2, and c = −2.

Similarly, Figure 8c shows the unconditional and conditional PDFs of $y$ for Example 7, and Figure 8d shows the corresponding conditional variance of $y$. In this case, fixing $x_1$ at any value between −0.58 and 0.58 causes $V(y|x_1)$ to be smaller than $V(y)$, whereas fixing $x_1$ at any value beyond this range causes $V(y|x_1)$ to be larger than $V(y)$. This means knowledge of the true value of $x_1$ can either decrease or increase the total uncertainty, depending on what the value actually is. However, if we average across the full ranges of $x_1$ and $x_2$, we find that $V(y) = E[V(y|x_1)] = E[V(y|x_2)]$, so that all of the variance in $y$ can seem to be caused by the interaction effect.

#### 4.2.3. Example 8

Figure 9a shows a quadratic function augmented by a constant interaction term, resulting in a parabolic surface rotated around the $y$ axis due to the interaction term. The interaction effect is constant across the factor space (Figure 9g). Method one indicates no direct effect, the direct effects computed by the second method include the impact of the interaction coefficient, while the direct effects computed by the third method do not depend on the interaction coefficient. It is interesting to note that the interaction term, in most of these examples, *does* contribute to the direct effects.

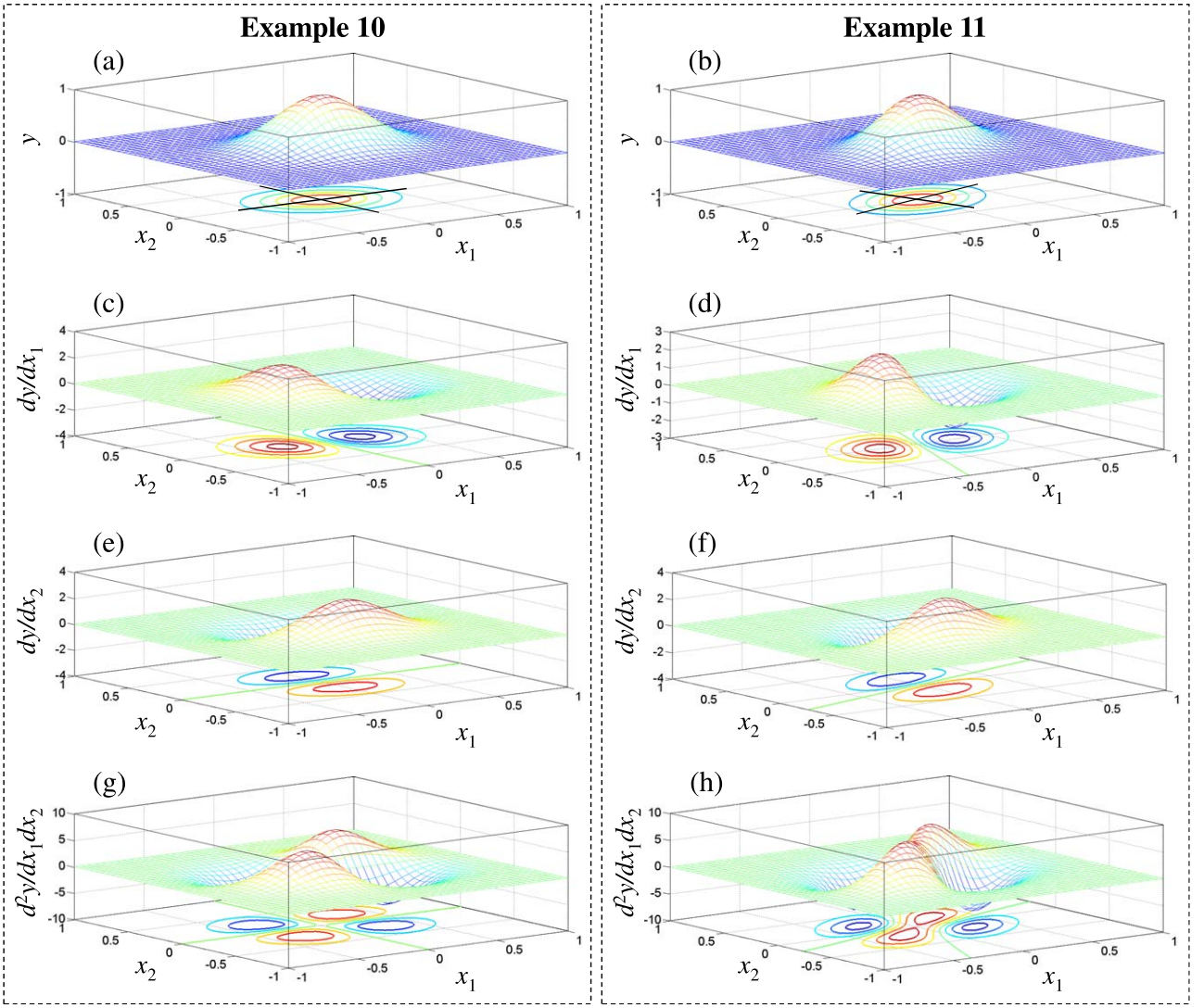


**Figure 10.** Examples 10 and 11 – Response surfaces in the forms of uncorrelated and correlated normal distributions: (a) and (b) response surfaces, (c) and (d) surfaces representing partial derivatives with respect to $x_1$, (e) and (f) surfaces representing partial derivatives with respect to $x_2$, (g) and (h) surfaces representing second-order derivatives.

#### 4.2.4. Example 9

Figure 9b shows a quadratic function augmented by a *variable* (*non*constant) interaction term. The local interaction effect follows a linear function (see Figure 9h), and although nonzero across the factor space, its average value is zero. Method one assesses the direct effect of $x_1$ as being zero, while the direct effect of $x_2$ depends on c. With method two, the direct effects of both of the factors depend on c, whereas with method three, only the direct effect of $x_2$ depends on c.

#### 4.2.5. Examples 10 and 11

Figures 10a and 10b show two response surfaces having the forms of bivariate normal distributions with $\sigma^2_{x_1}$ and $\sigma^2_{x_2}$ equal to 0.075 and 0.05, respectively. Example 10 (Figure 10) has no factor correlation while Example 11 (Figure 10b) has a factor correlation coefficient of 0.33. The density functions of these normal distributions have been multiplied by 0.385 and 0.364, respectively, to scale the range of the $y$ response surfaces to [0,1]. Such response surfaces are commonly encountered in model calibration, particularly in the vicinity of the main region of attraction in the factor space. In the model calibration context, the interaction effect in Example 10 would be considered to be zero, while Example 11 corresponds to an interaction effect defined and quantified by the correlation coefficient of the normal distribution. However, counter to intuition, the three methods report interaction effects that are

significant and relatively similar for both cases; for example, the variance-based method reports 37% of the total variance in Example 10 and 40% of the total variance in Example 11 to be due to the interaction effect between the two factors.

These examples show that, even in simple cases where explicit analytical equations are available, the attribution and assessment of direct and interaction effects is not trivial, and that the (supposed) "interaction" terms in the equations can also contribute to the direct effects. As with the direct effects, the interaction effects can vary (linearly or nonlinearly, and even multimodally) across the factor space, further complicating the interpretation of interaction effects, particularly in high-dimensional problems.

## 5. Final Remarks: The Need for Comprehensive Characterization of "Global" Sensitivity

The literature reports a variety of SA approaches that characterize different intuitive understandings of "sensitivity" in the context of model development and application. Each approach focuses on somewhat different characteristic properties of the underlying response surfaces, leading to differences, even conflicts, in the assessment of sensitivity. To-date, the Sobol and Morris approaches, and their extensions, provide the most rigorous "model-free" approaches to global sensitivity analysis. However, their utility is limited as follows:

1. The variance-based Sobol approach is based entirely on characterizing the global variance of model response, and its decomposition into components associated with individual contributing factors. As such, it is unable to distinguish between response surface structures that have identical global variance of model response but different distributions and spatial organizations (response surface structures) of the model response and its derivatives. In general, any approach that relies only on the distributional moments of model responses completely ignores the structure of the underlying response surface.
2. The Morris approach and its extensions attempt to globally aggregate local sensitivity information (first-order partial derivatives) across the factor space. This approach can characterize the structure of an underlying response surface to only a limited extent. However, all implementations of this approach are prone to the scale issue, and the step size of the analysis can have a significant impact on the conclusions about underlying sensitivities.

Our analysis indicates that at least four important characteristics must be considered when investigating and interpreting the sensitivity of a response surface (e.g., a metric of model performance) to its parameters/factors:

a. Local sensitivities (i.e., first-order derivatives)

b. The global distribution of local sensitivities (characterized, for example, by mean and variance)

c. The global distribution of model responses (characterized, for example, by variance)

d. The structural organization of the response surface (including, for example, its shape, multimodality, and degree of nonsmoothness/roughness).

Through illustrative examples, we have shown that existing GSA approaches typically focus on only one or a few of these characteristics while ignoring others. It seems evident that a comprehensive approach to GSA is required that can incorporate the information represented by all the above-mentioned characteristics within a single framework. In a separate paper, we propose such a framework that seeks to encapsulate all of these characteristics into a unified assessment of local and global sensitivity.

We have also discussed and illustrated the nontrivial nature of the problem of quantifying and interpreting interaction effects between different factors, even for low-dimensional problems. Analogous to local sensitivity, the local interaction effect can be quantified and interpreted through second and higher-order derivatives. However, the generalization of such local effects to global measures of interaction is not straightforward, and the variance-based approach remains the *only* existing approach that provides a meaningful global measure of interaction effects. In spite of this, the interpretation of interaction effects is not always easy, even in simple two-factor problems, and variance-based assessments can run counter to intuition.

Note that this paper is not intended to argue that existing GSA methods are not useful. Rather, we believe it is necessary to improve our understanding of the functionality of different existing methods, and to deepen the

conversation of what we, as the modeling community, intend when we conduct an analysis of sensitivity. This paper summarizes the fundamental differences between existing GSA methods, and helps explain why many published numerical comparisons of different methods have reported inconsistent and even conflicting assessments of sensitivity. We hope our analysis will provoke further conversation and research toward an improved understanding of sensitivity in the important context of EESMs development and application.

Overall, given the growing importance of sensitivity analysis to Earth and Environmental Systems modeling, it seems advisable to develop and utilize GSA techniques that effectively quantify a meaningful/intuitive sense of the manner in which a model response varies with changes in individual factors, and which incorporates all relevant sensitivity-related characteristics of a response surface. Further, there is a need to better understand what kinds of (model performance) response metrics are able to properly characterize the important aspects of behavior of an EESM (i.e., the model output response trajectory may be quite sensitive to a factor, but poorly chosen metrics can dampen or even remove this information). Only when these issues are properly addressed for the univariate case (single model response) can we consider the further, more complex, multivariate case involving multiple model responses (i.e., as in multiflux/multicriteria model calibration). As always, we invite discussion and collaboration with others interested in these and related issues of system identification and model development.

**Acknowledgments**
We would like to thank the Editor, Alberto Montanari, and Tony Jakeman and three anonymous reviewers for their comments that improved the quality of the paper. The first author thanks the support by the Canada Excellence Research Chair in Water Security, Howard Wheater, and the Global Institute for Water Security. The second author acknowledges support by the Australian Centre of Excellence for Climate System Science (CE110001028), and from the EU-funded project "Sustainable Water Action (SWAN): Building Research Links Between EU and US" (INCO-20011-7.6 grant 294947). The data used to support this paper (11 conceptual examples) are hypothetical and fully included in relevant sections of the paper. Interested readers may contact the first author at saman.razavi@usask.ca for more information.

## References

Bard, Y. (1974), *Nonlinear Parameter Estimation*, Academic, N. Y.

Bastidas, L. A., T. S. Hogue, S. Sorooshian, H. V. Gupta, and W. J. Shuttleworth (2006), Parameter sensitivity analysis for different complexity land surface models using multicriteria methods, *J. Geophys. Res.*, *111*, D20101, doi:10.1029/2005JD006377.

Campolongo, F., J. Cariboni, and A. Saltelli (2007), An effective screening design for sensitivity analysis of large models, *Environ. Modell. Software*, *22*(10), 1509–1518.

Chu-Agor, M. L., R. Muñoz-Carpena, G. Kiker, A. Emanuelsson, and I. Linkov (2011), Exploring vulnerability of coastal habitats to sea level rise through global sensitivity and uncertainty analyses, *Environ. Modell. Software*, *26*(5), 593–604.

Clark, M. P., D. Kavetski, and F. Fenicia (2011), Pursuing the method of multiple working hypotheses for hydrological modeling, *Water Resour. Res.*, *47*, W09301, doi:10.1029/2010WR009827.

Crosetto, M., S. Tarantola, and A. Saltelli (2000), Sensitivity and uncertainty analysis in spatial modelling based on GIS, *Agric. Ecosyst. Environ.*, *81*(1), 71–79.

Cukier, R. I., C. M. Fortuin, K. E. Shuler, A. G. Petschek, and J. H. Schaibly (1973), Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory, *J. Chem. Phys.*, *59*(8), 3873–3878.

Degenring, D., C. Froemel, G. Dikta, and R. Takors (2004), Sensitivity analysis for the reduction of complex metabolism models, *J. Process Control*, *14*(7), 729–745.

Duan, Q., S. Sorooshian, and V. Gupta (1992), Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resour. Res.*, *28*(4), 1015–1031.

Edwards, A. W. F. (1972), *Likelihood: An Account of the Statistical Concept of Likelihood and Its Application to Scientific Inference*, Cambridge Univ. Press, Cambridge, U. K.

Fisher, R. A. (1951), *The Design of Experiments*. Edinburgh: Oliver & Boyd.

Gupta, H. V., S. Sorooshian, and P. O. Yapo (1998), Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, *Water Resour. Res.*, *34*(4), 751–763.

Gupta, H. V., L. A. Bastidas, S. Sorooshian, W. J. Shuttleworth, and Z. L. Yang (1999), Parameter estimation of a land surface scheme using multicriteria methods, *J. Geophys. Res.*, *104*(D16), 19,491–19,503.

Gupta, H. V., T. Wagener, and Y. Q. Liu (2008), Reconciling theory with observations: Elements of a diagnostic approach to model evaluation, *Hydrol. Processes*, *22*(18), 3802–3813.

Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez (2009), Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, *J. Hydrol.*, *377*(1-2), 80–91.

Hamby, D. (1994), A review of techniques for parameter sensitivity analysis of environmental models, *Environ. Monit. Assess.*, *32*(2), 135–154.

Homma, T., and A. Saltelli (1996), Importance measures in global sensitivity analysis of nonlinear models, *Reliability Eng. Syst. Safety*, *52*(1), 1–17.

Iman, R. L., and W. J. Conover (1979), The use of the rank transform in regression, *Technometrics*, *21*(4), 499–509.

Kavetski, D., and M. P. Clark (2010), Ancient numerical daemons of conceptual hydrological modeling: 2. Impact of time stepping schemes on model analysis and prediction, *Water Resour. Res.*, *46*, W10511, doi:10.1029/2009WR008896.

Kleijnen, J. P. C. (1995), Sensitivity analysis and optimization of system dynamics models: Regression analysis and statistical design of experiments, *Syst. Dyn. Rev.*, *11*(4), 275–288.

Kleijnen, J. P. C. (2005), An overview of the design and analysis of simulation experiments for sensitivity analysis, *Eur. J. Oper. Res.*, *164*(2), 287–300.

Krzykacz-Hausmann, B. (2001), Epistemic sensitivity analysis based on the concept of entropy, in *Proceedings of SAMO2001*, edited by P. Prado and R. Bolado, pp. 31–35, CIEMAT, Madrid.

Kuczera, G., and E. Parent (1998), Monte Carlo assessment of parameter uncertainty in conceptual catchment models: The Metropolis algorithm, *J. Hydrol.*, *211*(1), 69–85.

Lidén, R., and J. Harlin (2000), Analysis of conceptual rainfall–runoff modelling performance in different climates, *J. Hydrol.*, *238*(3–4), 231–247.

Liu, H., W. Chen, and A. Sudjianto (2006), Relative entropy based method for probabilistic sensitivity analysis in engineering design, *J. Mech. Design*, *128*(2), 326–336.

Montgomery, D. C. (2008), *Design and Analysis of Experiments*, John Wiley & Sons, Inc., Hoboken, N. J.
Morris, M. D. (1991), Factorial sampling plans for preliminary computational experiments, *Technometrics*, *33*(2), 161–174.
Muleta, M. K., and J. W. Nicklow (2005), Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model, *J. Hydrol.*, *306*(1–4), 127–145.
Myers, R. H., D. C. Montgomery, and C. M. Anderson-Cook (2011), *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, John Wiley & Sons. Hoboken, N. J.
Nossent, J., P. Elsen, and W. Bauwens (2011), Sobol' sensitivity analysis of a complex environmental model, *Environ. Modell. Software*, *26*(12), 1515–1525.
Pappenberger, F., K. J. Beven, M. Ratto, and P. Matgen (2008), Multi-method global sensitivity analysis of flood inundation models, *Adv. Water Resour.*, *31*(1), 1–14.
Pietroniro, A., V. Fortin, N. Kouwen, C. Neal, R. Turcotte, B. Davison, D. Verseghy, E. Soulis, R. Caldwell, and N. Evora (2007), Development of the MESH modelling system for hydrological ensemble forecasting of the Laurentian Great Lakes at the regional scale, *Hydrol. Earth Syst. Sci.*, *11*(4), 1279–1294.
Rabitz, H., M. Kramer, and D. Dacol (1983), Sensitivity analysis in chemical-kinetics, *Annu. Rev. Phys. Chem.*, *34*, 419–461.
Rakovec, O., M. C. Hill, M. P. Clark, A. H. Weerts, A. J. Teuling, and R. Uijlenhoet (2014), Distributed Evaluation of Local Sensitivity Analysis (DELSA), with application to hydrologic models, *Water Resour. Res.*, *50*, 409–426, doi:10.1002/2013WR014063.
Ratto, M., S. Tarantola, and A. Saltelli (2001), Sensitivity analysis in model calibration: GSA-GLUE approach, *Comput. Phys. Commun.*, *136*(3), 212–224.
Razavi, S., B. A. Tolson, and D. H. Burn (2012a), Review of surrogate modeling in water resources, *Water Resour. Res.*, *48*, W07401, doi: 10.1029/2011WR011527.
Razavi, S., B. A. Tolson, and D. H. Burn (2012b), Numerical assessment of metamodelling strategies in computationally intensive optimization, *Environ. Modell. Software*, *34*, 67–86.
Rosolem, R., H. V. Gupta, W. J. Shuttleworth, X. Zeng, and L. G. G. Gonçalves (2012), A fully multiple-criteria implementation of the Sobol' method for parameter sensitivity analysis, *J. Geophys. Res.*, *117*, D07103, doi:10.1029/2011JD016355.
Ruano, M. V., J. Ribes, A. Seco, and J. Ferrer (2012), An improved sampling strategy based on trajectory design for application of the *Morris* method to systems with many input factors, *Environ. Modell. Software*, *37*, 103–109.
Sacks, J., W. J. Welch, T. J. Mitchell, and H. P. Wynn (1989), Design and analysis of computer experiments, *Stat. Sci.*, *4*(4), 409–435.
Saltelli, A., and P. Annoni (2010), How to avoid a perfunctory sensitivity analysis, *Environ. Modell. Software*, *25*(12), 1508–1517.
Saltelli, A., and R. Bolado (1998), An alternative way to compute Fourier amplitude sensitivity test (FAST), *Comput. Stat. Data Anal.*, *26*(4), 445–460.
Saltelli, A., S. Tarantola, and K.-S. Chan (1999), A quantitative model-independent method for global sensitivity analysis of model output, *Technometrics*, *41*(1), 39–56.
Saltelli, A., S. Tarantola, F. Campolongo, and M. Ratto (2004), *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*, John Wiley.
Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola (2008), *Global Sensitivity Analysis: The Primer*, John Wiley & Sons.
Shin, M.-J., J. H. Guillaume, B. F. Croke, and A. J. Jakeman (2013), Addressing ten questions about conceptual rainfall–runoff models with global sensitivity analyses in R, *J. Hydrol.*, *503*, 135–152.
Sobol', I. M. (2001), Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Math. Comput. Simul.*, *55*(1-3), 271–280.
Sobol, I. M., and S. Kucherenko (2009), Derivative based global sensitivity measures and their link with global sensitivity indices, *Math. Comput. Simul.*, *79*(10), 3009–3017.
Sobol', I. M. (1990), Sensitivity estimates for nonlinear mathematical models [in Russian], *Mat. Modelirovanie*, *2*(1), 112–118.
Spear, R. C., and G. M. Hornberger (1980), Eutrophication in peel inlet—II. Identification of critical uncertainties via generalized sensitivity analysis, *Water Res.*, *14*(1), 43–49.
Spear, R. C., T. M. Grieb, and N. Shang (1994), Parameter uncertainty and interaction in complex environmental models, *Water Resour. Res.*, *30*(11), 3159–3169.
van Griensven, A., T. Meixner, S. Grunwald, T. Bishop, M. Diluzio, and R. Srinivasan (2006), A global sensitivity analysis tool for the parameters of multi-variable catchment models, *J. Hydrol.*, *324*(1–4), 10–23.
Vrugt, J. A., H. V. Gupta, W. Bouten, and S. Sorooshian (2003), A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters, *Water Resour. Res.*, *39*(8), 1201, doi:10.1029/2002WR001642.