

Water Resources Research

RESEARCH ARTICLE

10.1029/2019WR025436

Key Points:

- We extend the theory of VARS to handle non-uniformly distributed and correlated inputs
- We develop a generalized sampling strategy for VARS to handle analytical multivariate distributions
- We show that proper accounting for correlation effects, which are often ignored, is essential in sensitivity and uncertainty analysis

Supporting Information:

- Supporting Information S1

Correspondence to:

N. C. Do,
nchu.do@usask.ca

Citation:

Do, N. C., & Razavi, S. (2020). Correlation effects? A major but often neglected component in sensitivity and uncertainty analysis. *Water Resources Research*, 56, e2019WR025436. <https://doi.org/10.1029/2019WR025436>

Received 25 APR 2019

Accepted 23 FEB 2020

Accepted article online 25 FEB 2020

Correlation Effects? A Major but Often Neglected Component in Sensitivity and Uncertainty Analysis

Nhu Cuong Do^{1,2,3}  and Saman Razavi⁴ 

¹Global Institute for Water Security, University of Saskatchewan, Saskatoon, Saskatchewan, Canada, ²School of Civil, Environmental and Mining Engineering, The University of Adelaide, Adelaide, South Australia, Australia, ³Department of Construction Management, Thuy Loi University, Hanoi, Vietnam, ⁴Global Institute for Water Security and School of Environment and Sustainability, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

Abstract Global sensitivity analysis (GSA) provides essential insights into the behavior of Earth and environmental systems models and identifies dominant controls of output uncertainty. Previous work on GSA, however, has typically been under the assumption that the controlling factors such as model inputs and parameters are independent, whereas, in many cases, they are correlated and their joint distribution follows a variety of forms. Although this assumption can limit the credibility of GSA and its results, very few studies in the field of water and environmental modeling address this issue. In this paper, we first discuss the significance of correlation effects in GSA and then propose a new GSA framework for properly accounting for correlations in input/parameter spaces. To this end, we extend the “variogram-based” theory of GSA, called variogram analysis of response surfaces (VARS), and develop a new generalized star sampling technique (called gSTAR) to accommodate correlated multivariate distributions. We test the new gSTAR-VARS method on two test functions, against a state-of-the-art GSA method that handles correlation effects. We then apply gSTAR-VARS to the HBV-SASK model, calibrated via a Bayesian, Markov chain Monte Carlo approach, for design flood estimation in the Oldman River Basin in Canada. Results demonstrate that accounting for correlation effects can be critically important in GSA, especially in the presence of nonlinearity and interaction effects in the underlying response surfaces. The proposed method can efficiently handle correlations and different distribution types and simultaneously generate a range of sensitivity indices, such as total-variogram effects, variance-based total-order effects, and derivative-based elementary effects.

1. Introduction

Environmental research requires multidisciplinary knowledge to explore, explain, and predict the reaction of natural systems to changing conditions. Such important investigations typically include the application of simulation models. Various types of models, ranging from a simple rainfall-runoff process formulation to complex dynamical Earth system models (DESMs), have been developed to fulfill specific tasks and research objectives. These models often involve a number of hypotheses and parameters to estimate one or a number of output variables of interest. However, the hypotheses and parameter values are subject to variation and uncertainty, which results in variation and uncertainty in the model outputs. Consequently, there is a need to evaluate the role and impact of uncertainty in a range of “factors” (i.e., input variables and parameters) on the uncertainty in model outputs.

Sensitivity analysis (SA) is an effective tool for gaining insights into model behavior, evaluating the importance of different model factors, and identifying the dominant controls of output uncertainty. The objectives for the implementation of SA have been comprehensively reviewed in a number of papers, such as Saltelli et al. (2006), Razavi and Gupta (2015), Song et al. (2015), Pianosi et al. (2016), Gupta and Razavi (2018), and Razavi et al. (2019). These include diagnostic testing and assessment of similarity (e.g., Gupta et al., 2008; Saltelli et al., 2004), identification of factor importance and function (e.g., Haghnegahdar et al., 2017; Muleta & Nicklow, 2005; Razavi & Gupta, 2016b), characterization of regions in the input space with different sensitivity properties (e.g., Do et al., 2016; Rakovec et al., 2014), analysis of factor interdependence (e.g., Lidén & Harlin, 2000; Nossent et al., 2011), identification of noninfluential factors (e.g., Do et al., 2018;

van Griensven et al., 2006), and uncertainty apportionment (e.g., Chu-Agor et al., 2011). Depending on their scope and applicability, SA methods can be classified into different categories, the most common of which are local sensitivity analysis (LSA) and global sensitivity analysis (GSA).

LSA methods are used to explore the response surface in the neighborhood of a specified set of factors. These methods are generally based on partial derivatives of a model output functions with respect to factors of interest at some nominal locations in the factor space. Because of their relatively simple mechanism, easy implementation, and low computational cost, LSA methods have been widely used across an extensive range of modeling disciplines. LSA results, however, are only valid at the nominal locations (Saltelli & Annoni, 2010) and may be dramatically different at different parts of the factor space, especially in case of nonlinear models (Rakovec et al., 2014).

GSA methods, on the other hand, examine the sensitivity of model outputs to model factors by taking samples across their entire feasible factor space. Various techniques can be used to perform a GSA, such as regional sensitivity analysis (Spear & Hornberger, 1980; Wagener et al., 2001), variance-based methods (Homma & Saltelli, 1996; Saltelli et al., 2006; Sobol, 1993), moment-independent methods (Pianosi & Wagener, 2018; Plischke et al., 2013), derivative-based methods (globally aggregated measures of local sensitivities) (Morris, 1991), and variogram-based methods (Razavi & Gupta, 2016a, 2016b). Although commonly applied to environmental models, two major issues arise with GSA methods, as highlighted by Razavi and Gupta (2015). First, different GSA methods are developed based on different theoretical definitions of sensitivity, which may lead to different, even conflicting, sensitivity results for a given problem. Second, the computational cost of implementing GSA for a DESM can be large for computationally intensive and/or high-dimensional models. This cost may limit the applicability of several GSA methods due to their computational inefficiency, which are often assessed in terms of the number of samples (model simulation runs) required to provide reliable and robust results.

Razavi and Gupta (2016a) introduced an approach to GSA, called variogram analysis of response surfaces (VARS), which provides an alternative to cope with these aforementioned issues. The theoretical relationship between VARS and other GSA approaches allows VARS to compute the Sobol's (variance-based) and Morris's (derivative-based) sensitivity measures without extra computational expense. The first numerical implementation of VARS has been via the recently developed STAR sampling strategy (Razavi & Gupta, 2016b). This STAR-VARS framework is highly efficient and, together with a factor grouping strategy, can handle very high dimensional problems (Sheikholeslami et al., 2019). STAR-VARS also employs bootstrapping (Razavi & Gupta, 2016b), which together facilitate rapid convergence and provide a measure of uncertainty in the results.

Like many other GSA methods, the current STAR-VARS framework can only be applied to models whose factors are independent and uniformly distributed. This should in principle limit the applicability of VARS, as many environmental and hydrological models involve factors (i.e., inputs, parameters, and boundary conditions) whose true/estimated distribution is essentially nonuniform and/or correlated with others (see section 2). Therefore, in this paper, we extend the theory of VARS to accommodate correlated and/or nonuniformly distributed factors and demonstrate its applicability and benefit to the analysis of complex environmental models. At a more fundamental level, we argue that there is a general ignorance of "correlation effects" in many sensitivity and uncertainty analysis studies. As such, through illustrative examples, we also aim to provide insight into the role and significance of correlation among factors and demonstrate the need for using proper methods that carefully and adequately address possible correlation effects in sensitivity and uncertainty analyses.

The paper is structured as follows. In section 2, we highlight the existence and significance of correlation and dependency between factors in DESMs and review existing GSA methods that can handle correlation effects in the field of environmental and water resources modeling. This is followed in section 3 by an overview of the variogram-based GSA and VARS theory. In section 4, we formulate a new version of VARS that accounts for correlation effects and describe its implementation process. The proposed framework is then illustrated and evaluated on two "test function" examples and a real case study in sections 5–7. This real-world case study is an SA of design floods obtained from the HBV-SASK hydrologic model to its parameters in the Oldman River Basin, Canada. Conclusions and suggestions for future work are given in section 8.

2. The Quest for GSA With Correlated Factors

The term *correlation* refers to the statistical dependency among pairs of factors. This is different from *interaction*, which refers to the presence of nonadditive operations between two or more factors within a model's structural equations. Correlations can be statistically derived from samples collected in a field or lab experiment, independent of a numerical model (e.g., Cosby et al., 1984; Lemke et al., 2004; Pan et al., 2011). Or they can be inferred for a particular numerical model via the process of model inversion when sufficient observed data are available. In the latter case, the derived correlations depend on the model structure, constitutive equations, and the likelihood function used for inversion and, therefore, are influenced by interactions. An example tool to calibrate model parameters and infer their correlations a posteriori of model inversion is the Differential Evolution Adaptive Metropolis (DREAM) algorithm (Vrugt, 2016), which is a Bayesian approach of model calibration/inversion. Other methods can identify parameter correlations by analyses of the model Hessian matrix (e.g., Chu & Hahn, 2007).

Although correlation is a relatively common feature among factors involved in environmental and water systems-related models, its effects are often ignored when carrying out sensitivity and uncertainty analysis, perhaps for mathematical and technical convenience. This is despite the fact that such correlations can significantly control model outputs and their associated uncertainty, as shown in the following example.

2.1. Effect of Correlations on Variance-Based Sensitivity Indices—An Illustrative Example

Here we use a simple additive example adapted from Xu and Gertner (2008) to illustrate the effect of correlations and the complexities that its presence gives rise to when estimating sensitivity measures. In this example, $Y = 2X_1 + 3X_2$, where X_1 and X_2 are standard, normally distributed random variables, that is, $X_1 \sim N(\mu_1 = 0, \sigma_1 = 1)$ and $X_2 \sim N(\mu_2 = 0, \sigma_2 = 1)$, with a Pearson correlation coefficient of ρ . It can be shown that the variance-based first-order sensitivity indices are $S_1 = \frac{V[E(Y|X_1)]}{V(Y)} = \frac{(2+3\rho)^2}{13+12\rho}$ and $S_2 = \frac{V[E(Y|X_2)]}{V(Y)} = \frac{(3+2\rho)^2}{13+12\rho}$, which are nonlinear functions of the correlation between the two factors (Figure 1a). As shown, ignoring correlation (i.e., assuming $\rho = 0$) may result in sensitivity indices far from the real values when the correlation significantly deviates from 0.

Figure 1(b) shows the variance-based total-order indices of X_1 and X_2 , which are $ST_1 = \frac{E[V(Y|X_2)]}{V(Y)} = \frac{4(1-\rho^2)}{13+12\rho}$ and $ST_2 = \frac{E[V(Y|X_1)]}{V(Y)} = \frac{9(1-\rho^2)}{13+12\rho}$, for different values of the correlation coefficient. As expected, the total-order indices also change as ρ varies, and interestingly, for positive values of ρ , the total-order indices become smaller than the first-order indices. Figures 1c and 1d show ratios of factor sensitivity, defined as the sensitivity index of a factor divided by the sum of the sensitivity indices for all factors, based on the first-order and total-order effects, respectively. Accordingly, the ratios of factor sensitivity based on the first-order indices change for different values of ρ , whereas, perhaps counterintuitively, the same figures based on the total-order indices remain unchanged. Detailed derivation of the above equations is presented in the supporting document.

We note that all the above equations were derived analytically, because of the simplicity of the test function and the assumed multivariate normal distribution for the factors. However, in real-world problems, where the multivariate distribution of factors is more complex and the model response surface is nonlinear and nonadditive, GSA must be either performed numerically via sampling or used with some simplified surrogate functions. The next subsection provides a review of the existing GSA methods designed for this purpose.

2.2. Review of GSA Methods for Correlated Factors

Several GSA methods in the literature have been modified to deal with correlated factors. These GSA methods can be classified into either parametric, that is, model-based methods (e.g., Da Veiga et al., 2009; Li et al., 2010; Mara & Tarantola, 2012; Xu & Gertner, 2008), or nonparametric, that is, model-free methods (e.g., Kucherenko et al., 2012; Mara et al., 2015; Tarantola & Mara, 2017).

Parametric GSA methods assume a priori a model form for the underlying response surface and often develop an approximation of it, commonly referred to as a "metamodel" or "surrogate model" (Razavi et al., 2012). For example, Xu and Gertner (2008) proposed a regression-based method for estimating the variance-based sensitivity of model outputs to correlated factors, in which the total sensitivities are divided into correlated and uncorrelated contributions. This method, however, is only applicable to models with

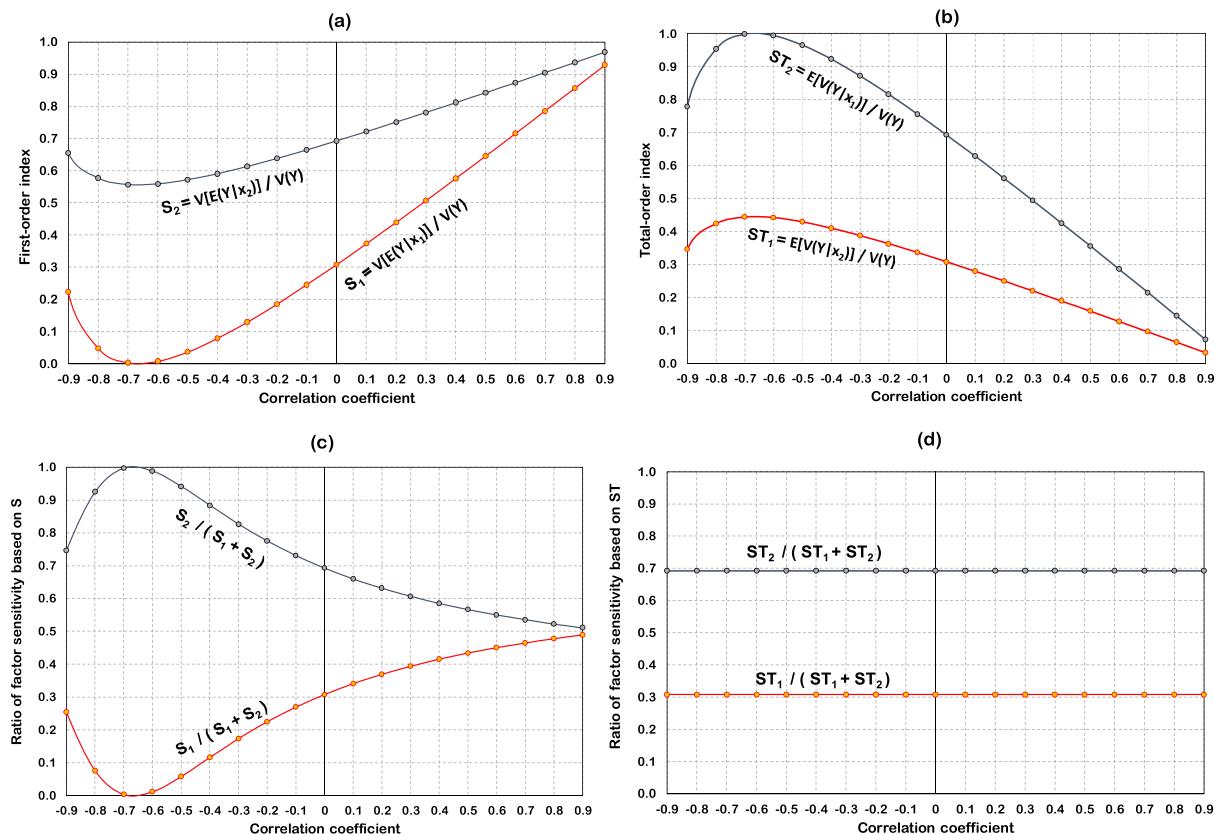


Figure 1. The variance-based (a) first-order and (b) total-order sensitivity indices corresponding to different correlation coefficient values of the illustrative example $Y = 2X_1 + 3X_2$. Ratios of factor sensitivity based on the (c) first-order and (d) total-order sensitivity indices.

linear or approximately linear response surfaces. Da Veiga et al. (2009) introduced a method using local polynomial estimators of the conditional expectation and variance of the model output given each factor to estimate variance-based sensitivity indices. Li et al. (2010) decomposed the total variance of the model output into the covariance of expansion functions (i.e., the piecewise polynomial functions) and estimated the sensitivity indices based on this decomposition. Mara and Tarantola (2012) introduced a modified variance-based method to compute the correlated and uncorrelated contributions to the total variance of model output via a polynomial chaos expansion.

Nonparametric GSA methods, on the other hand, are deemed more suitable in the context of computer-based modeling, because the form and complexity of the response surface of a computer model cannot be known *a priori* (Razavi & Gupta, 2015). To deal with the correlation among factors, these methods generally involve transformation techniques for sampling from multivariate distributions of the correlated factors. For example, Kucherenko et al. (2012) used copula transformations to extend the variance-based method to handle correlated factors. The Rosenblatt transformation was selected in Mara et al. (2015), while the Rosenblatt transformation and the Nataf transformation were proposed in Tarantola and Mara (2017).

While such methods to deal with correlated factors exist, very few GSA studies in the field of environmental and water-related modeling consider the correlations among factors; examples include Manache and Melching (2008), Pan et al. (2011), and Zhu (2012), which use parametric methods, and Ahn (1996), which uses a nonparametric method. These studies, however, simplified the correlation problem by assuming the factors are either linearly tied together or normally distributed to take advantage of the convenient characteristics of the normal distribution. While their assumptions may be justified for specific case studies, their approaches may not be applicable to broader cases where the factors are, for example, nonnormally distributed. This is the case of complex environmental models, where much research is yet to be done to properly

investigate and account for correlation effects. Therefore, this paper provides an efficient alternative using the theory of VARS to properly account for nonuniform distributions and correlation effects among factors in GSA.

3. Overview of the Variogram-Based SA

VARS is a relatively new theory for GSA (Razavi & Gupta, 2016a, 2016b), in which directional variogram and covariogram functions are used to characterize sensitivity information around the factors of interest. Suppose that the response surface Y of a model is represented by a function:

$$Y = f(X_1, X_2, \dots, X_n), \quad (1)$$

where X_1, X_2, \dots, X_n are n factors (e.g., model inputs or parameters) to be considered in SA. The sensitivity of Y with respect to factor X_i is a “perturbation scale-dependent” property that can be obtained by computing the variogram γ and covariogram C of Y in the direction of factor X_i :

$$\gamma(h_i) = \frac{1}{2} V\{f(X'_i, X_{\sim i}) - f(X_i, X_{\sim i})\}, \quad (2)$$

$$C(h_i) = Cov\{f(X'_i, X_{\sim i}), f(X_i, X_{\sim i})\}, \quad (3)$$

where $X'_i = X_i + h_i$; h_i is the perturbation size in X_i (i.e., distance between two points in the factor range); $V(\cdot)$ and $Cov(\cdot)$ respectively denote the variance and covariance operators; and $X_{\sim i}$ represents the set of all factors except X_i . Equation (2), under the “constant mean assumption,” can be written as

$$\gamma(h_i) = \frac{1}{2} E\left\{[f(X'_i, X_{\sim i}) - f(X_i, X_{\sim i})]^2\right\}, \quad (4)$$

where $E(\cdot)$ is the expectation operator.

Razavi and Gupta (2016a) introduced a new set of sensitivity indices called IVARS (i.e., integrated variogram across a range of scales) that comprehensively measure sensitivities across a range of perturbation scales by integrating the variogram function to a particular perturbation scale (H_i): $\Gamma(H_i) = \int_0^{H_i} \gamma(h_i) dh_i$. The IVARS₁₀, IVARS₃₀, and IVARS₅₀ values, corresponding to 10%, 30%, and 50% of the factors’ range, respectively, are recommended for use as the sensitivity measures (although any value from 0% to 50% of the range could be computed as a sensitivity measure). IVARS₅₀, called “total-variogram effect,” is deemed the most comprehensive index of global sensitivity, because it integrates sensitivity information across the full range of perturbation scales. Razavi and Gupta (2016b) proposed the STAR-VARS method to estimate the directional variograms, IVARS indices, and the Morris and Sobol-based indices. The steps required by the STAR-VARS method can be summarized as follows: (1) choose the “resolution” of sampling Δh , which is the smallest value for h for every factor ($\Delta h = 0.1$, i.e., 10% of the factor range, typically works well); (2) generate N_{Star} star centers in the factor space; and (3) for each star center, generate cross sections, that is, sample equally spaced points, Δh apart, along the direction of each factor passing through that star center; (4) collect all of the sample pairs at each certain distance (e.g., $h = 0.1$ and 0.2) and compute the directional variograms and covariograms; and finally, (5) numerically integrate the obtained variograms to the scale(s) of interest H_i to compute the IVARS indices. The derived covariograms are used to estimate Sobol variance-based indices as explained in Razavi and Gupta (2016b). We note that the constant mean assumption does not typically hold for complex response surfaces. Its violation, however, does not affect the IVARS indices, as they are directly defined based on equation (4). This assumption is basically needed to provide the theoretical relationship between the variance-based and variogram-based GSA approaches, as described in Razavi and Gupta (2016a), and therefore, its violation can have an impact on the VARS-based estimates of “total-order effects” (termed VARS-TO). As discussed and shown via an example in section 4.1.5 of Razavi and Gupta (2016b), this impact becomes marginal when Δh is small.

As mentioned in section 1, the STAR-VARS method developed in Razavi and Gupta (2016b) is for problems where factors are independent and uniformly distributed. In the next section, we extend the VARS theory to

accommodate any multivariate analytical distribution with correlation effects and develop a generalized STAR sampling strategy, called gSTAR, to enable application of VARS to any problems with nonuniform and correlated factors.

4. VARS Framework for Models With Correlated Factors

4.1. Theoretical Basics

For a set $X = \{X_1, X_2, \dots, X_n\}$ of correlated factors following a multivariate distribution $p(X_i, X_{-i})$, the variogram function in the direction of X_i (i.e. equation (4)) can be rewritten as

$$\gamma(h_i) = \frac{1}{2} E \left\{ \left[f\left(X'_i|_{X_{-i}}, X_{-i}\right) - f\left(X_i|_{X_{-i}}, X_{-i}\right) \right]^2 \right\}, \quad (5a)$$

where X_{-i} , $X_i|_{X_{-i}}$, and $X'_i|_{X_{-i}}$ follow probability distributions $p(X_{-i})$, $p\left(X_i|_{X_{-i}}\right)$, $p\left(X'_i|_{X_{-i}}\right)$, respectively. Therefore, equation (5a) can be further expressed as

$$\gamma(h_i) = \frac{1}{2} \int_{\Omega^{n+1}} \left[f\left(X'_i|_{X_{-i}}, X_{-i}\right) - f\left(X_i|_{X_{-i}}, X_{-i}\right) \right]^2 p\left(X'_i|_{X_{-i}}\right) p\left(X_i|_{X_{-i}}\right) p(X_{-i}) dX'_i dX_i dX_{-i}, \quad (5b)$$

where Ω^{n+1} is the $(n + 1)$ -dimensional domain of integration. Based on the law of conditional probability, we have $p\left(X_i|_{X_{-i}}\right)p(X_{-i}) = p(X_i, X_{-i})$. Therefore,

$$\gamma(h_i) = \frac{1}{2} \int_{\Omega^{n+1}} \left[f\left(X'_i|_{X_{-i}}, X_{-i}\right) - f\left(X_i|_{X_{-i}}, X_{-i}\right) \right]^2 p\left(X'_i|_{X_{-i}}\right) p(X_i, X_{-i}) dX'_i dX_i dX_{-i}. \quad (5c)$$

The above integral can be numerically estimated by sampling from $p(X_i, X_{-i})$ and $p\left(X'_i|_{X_{-i}}\right)$ using the following equation:

$$\hat{\gamma}(h_i) = \frac{1}{2N_h} \sum_{i=1}^{N_h} \left[f\left(X'_i|_{X_{-i}}, X_{-i}\right) - f\left(X_i|_{X_{-i}}, X_{-i}\right) \right]^2, \quad (6)$$

where N_h is the number of pairs corresponding to perturbation scale h , sampled along the i th direction.

4.2. Generalized Star-Based (gSTAR) Sampling

In this section, we introduce a new sampling strategy, called generalized star-based (gSTAR) sampling, which enables estimation of variogram and covariogram functions of response surfaces with nonuniform and correlated factors. For this purpose, marginal distributions of factors and their pairwise Pearson coefficients of correlation must be known.

Users of the gSTAR sampling strategy need to choose N_{star} and N_{cr} , which are the numbers of star centers and cross-sectional points, respectively. Like in STAR, choosing these numbers depends on the available computational budget, as the total number of sample points, and therefore model runs, is $N_{star}(1+n, N_{Cr})$, where n is the number of factors. Similar to any other sampling-based method, the larger the sample point, the more accurate the GSA result. In addition, loosely speaking, N_{star} needs to be larger than N_{cr} to better reproduce the correlation structure between the factors.

The gSTAR process follows four steps: (1) generate N_{star} star centers using the inverse Nataf transformation, which samples the joint probability distribution function $p(X_i, X_{-i})$; (2) for every star center, generate cross sections by sampling N_{cr} points along each direction X_i from the conditional distribution $p(X_i|X_{-i})$; (3) extract all pairs along each cross section; and (4) compute the directional variograms and covariograms. Details of each step are illustrated below using test function $Y = 1.11X_1^2 + X_2^2 - 0.2\cos(7\pi X_2)$, where X_1 ,

$X_2 \sim U(-1,1)$ and the correlation between X_1 and X_2 is -0.8 (example adopted from Razavi & Gupta, 2016a). This test function will also be used in section 5 as Example 1.

4.2.1. Generation of Star Centers Via Inverse Nataf Isoprobabilistic Transformation

Consider the marginal cumulative distribution functions (CDFs) of the factors to be $F_{X_1}(X_1), \dots, F_{X_n}(X_n)$ and the linear correlation coefficient between two factors X_i and X_j to be $\rho_{X_i,j}$; these correlation coefficients are the elements of the correlation matrix R_X . The generation of N_{star} star centers is carried out by application of the inverse Nataf transformation, which maps samples from a multivariate standard normal space (Z) to the “original factor space” of correlated factors (\mathbf{X}). To do so, we first estimate the respective “fictive” linear correlation matrix R_Z in the standard normal space of factors $\mathbf{Z} = \{Z_1, \dots, Z_n\}$ through the following relationship:

$$\rho_{X_i,j} = \frac{1}{\sigma_{X_i}\sigma_{X_j}} \iint_{\Omega^2} \left\{ \left[F_{X_i}^{-1}[\Phi(Z_i)] - \mu_{X_i} \right] \left[F_{X_j}^{-1}[\Phi(Z_j)] - \mu_{X_j} \right] \varphi_2(Z_i, Z_j, \rho_{Z_i,j}) dZ_i dZ_j \right\}, \quad (7)$$

where σ_{X_i} and σ_{X_j} are standard deviation values of factors X_i and X_j , respectively; $\rho_{Z_i,j}$ is the correlation coefficient between standard normal factors Z_i and Z_j ; Ω^2 is the two-dimensional (Z_i and Z_j) domain of integration; $\Phi(\cdot)$ represents the CDF of the respective standard normal factor; and φ_2 is the bivariate correlated standard normal probability density function. This equation is commonly solved by iterative algorithms, in which an initial guess of $\rho_{Z_i,j}$ is chosen and adjusted until satisfying equation (7). We note that the double integral in equation (7) may not have a solution for some certain pairs of marginal distributions. In that case, other similar distributions or the empirical equations suggested in Liu and Der Kiureghian (1986) can be used.

Once the fictive correlation matrix R_Z is estimated, N_{star} points are sampled from the standard normal space (\mathbf{Z}) with R_Z and transformed into the original factor space (\mathbf{X}) with R_X by T_1 :

$$T_1: \mathbf{Z}^{yields} \rightarrow \mathbf{X} = \begin{pmatrix} F_{X_1}^{-1}[\Phi(Z_1)] \\ F_{X_2}^{-1}[\Phi(Z_2)] \\ \vdots \\ F_{X_n}^{-1}[\Phi(Z_n)] \end{pmatrix}. \quad (8)$$

Figures 2a and 2b provide an illustrative example of the generation of star centers for the test function via the inverse Nataf isoprobabilistic transformation. In this example, $N_{star} = 500$ stars are plotted and the fictive correlation coefficient of the correlated bivariate standard normal distribution is estimated by equation (7) to be $\rho_{Z_{1,2}} = -0.8135$.

4.2.2. Generation of Cross Sections

For each star center generated, we need to design cross-sectional samples along the direction of every factor. First, for every cross section i passing through the star center k ($k = 1, \dots, N_{star}$) in the multivariate standard normal space, the mean $(\mu_{Z_i|Z_{\sim i}})_k$ and standard deviation $\Sigma_{Z_i|Z_{\sim i}}$ of the conditional probability distributions are calculated using basic multivariate statistics (see supporting information for details). Second, a point is randomly sampled from each of these cross sections following a normal distribution with the respective conditional mean and standard deviation. Third, these cross-sectional samples are mapped onto the original factor space (\mathbf{X}) using equation (8). The second and third steps above are repeated N_{cr} times, until all cross-sectional samples are collected.

For the example, Figures 2c and 2e show the cross sections of six of the star centers at the two directions in the multivariate standard normal space. The number of cross-sectional points (N_{cr}) chosen here is 10. Figures 2d and 2f show the same cross-sectional points mapped onto the original factor space. All the sampled points are subsequently evaluated by function f (i.e., a model).

4.2.3. Extract Pairs

Once the cross-sectional points are sampled, every possible pair of points falling on the same cross section is extracted, and the difference in the model responses of the two associated points, $\frac{1}{2} [f(X'_i|X_{\sim i}, X_{\sim i}) - f(X_i|X_{\sim i}, X_{\sim i})]^2$, is computed. The total number of pairs N_{pair} obtained from gSTAR is

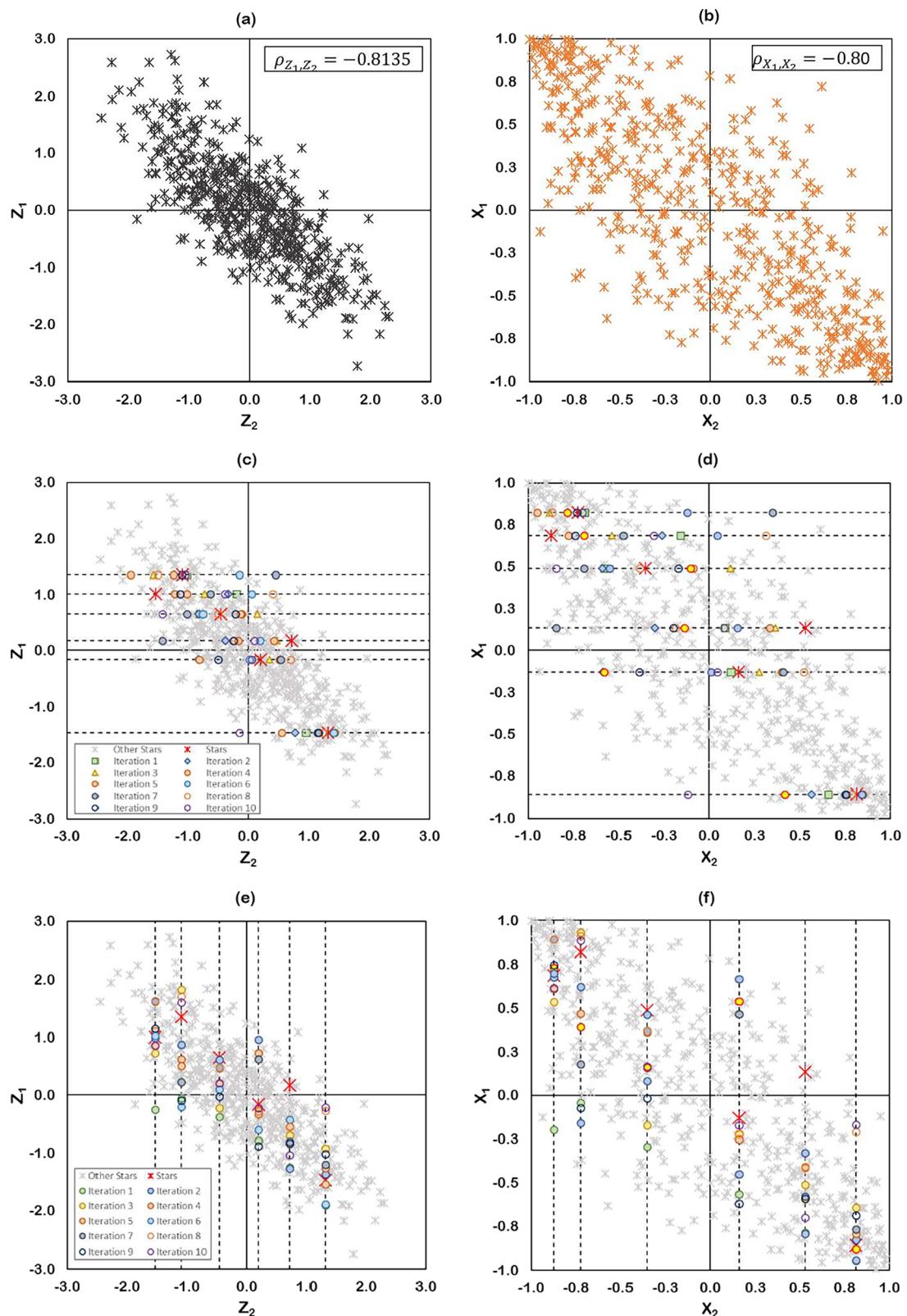


Figure 2. The generation of star centers: (a) star centers sampled in the standard normal space and (b) star centers in the original factor space, which is transformed from the standard space. The generation of cross sections of two factors, X_1 and X_2 , for six star centers, with each cross section having 10 samples: (c, e) cross sections in the standard normal space and (d, f) cross sections in the original variable space.

$$N_{pair} = \frac{N_{Cr}(N_{Cr}-1)}{2} * n * N_{star}. \quad (9)$$

4.2.4. Compute Cross-Sectional Variograms and Covariograms

Before computing variograms and covariograms, similar to the STAR sampling process, all the original factor ranges are scaled into the range [0,1]. Unlike in STAR, however, because the cross-sectional points are not equally spaced, no pair (m,l) is likely to exactly satisfy $|X_i(m) - X_i(l)| = \kappa \cdot \Delta h$, where Δh is any specific resolution and κ is a positive integer. Therefore, for each star and each direction, all the pairs are clustered into some number of intervals (i.e., bins), N_{bin} , which can be viewed as the reciprocal of the resolution Δh , defined in STAR. The directional variograms and covariograms are then computed using equation (6) based on the pairs collected in the bins. It should be noted that this binning technique is a standard practice in variogram analysis, because irregularly distributed data are common in applied geostatistics (examples can be found in SAS Institute, 1999, or Marchant & Lark, 2004). A proper N_{bin} must be chosen to ensure that there is a large enough number of pairs in each bin (e.g., Smith, 2016, suggests a minimum of 30 pairs for each bin).

For the illustrative example, Figures 3a and 3b plot the variogram clouds (squared markers), the estimated directional variograms and their 90% confidence intervals (dashed lines) using bootstrapping technique (Razavi & Gupta, 2016a, 2016b), and the “true” variograms (continuous lines). These estimated directional variograms and their confidence intervals were obtained using $N_{star} = 10$, $N_{cr} = 10$, and 20 bins (i.e., $\Delta h = 0.05$), while the true variograms were obtained using $N_{star} = 1000$, $N_{cr} = 500$, and 100 bins ($\Delta h = 0.01$). It is seen that, in this case, gSTAR has been able to estimate the directional variograms with a reasonable accuracy with a small number of sample points.

5. Illustrative Test Functions

In this section, we evaluate the proposed gSTAR-VARS method and analyze the correlation effects on sensitivity results using two test functions. The first example is adopted from Razavi and Gupta (2015, 2016a), where it was developed to demonstrate the effects of the spatial structure of the response surface on sensitivity results and that the variance-based method was unable to capture spatial structures. The second example is the Ishigami function, which has been used in various SA papers, such as Homma and Saltelli (1996), Kucherenko et al. (2012), and Plischke et al. (2013). This test function is considered a benchmark for GSA due to its nonlinear and nonmonotonic features and the presence of interaction effects. In these examples, Sobol's total-order effect indices are used as a benchmark for comparisons. In particular, we compare the gSTAR-VARS results to the numerically derived results by the method of Kucherenko et al. (2012), where the Sobol's method is modified to deal with correlated factors. The computer code of the Kucherenko et al. (2012) method was obtained from the authors and recoded to accommodate additional functionalities for application to our real case study in section 6. Our additions to the Kucherenko et al. (2012) method include the capability of handling asymmetrical triangular distributions and a more efficient strategy to generate conditional normal distributions for the calculation of the Sobol's indices. We also provide additional comparisons in the supporting information using the illustrative test function in section 2.1 based on analytically derived total-order indices. Those comparisons confirm that gSTAR-VARS is able to generate perfectly accurate results, matching with the analytical solution.

5.1. Example 1: $Y = 1.11X_1^2 + X_2^2 - 0.2\cos(7\pi X_2)$, where $X_1, X_2 \sim U(-1,1)$

This function is an additive combination of a unimodal function $Y = 1.11X_1^2$ and a multimodal function $Y = X_2^2 - 0.2\cos(7\pi X_2)$, as seen in Figure 4a. To show the effect of correlation on the results of GSA, we run gSTAR-VARS ($N_{star} = 1000$, $N_{cr} = 100$, and $\Delta h = 0.01$) for two cases: (1) X_1 and X_2 are uncorrelated and (2) X_1 and X_2 are negatively correlated with $\rho = -0.8$. Figures 4b and 4c plot the directional variograms along X_1 and X_2 directions, respectively. As expected, the variograms change due to the correlation effect; the variogram increases at smaller perturbation scales (e.g., at $h \leq 0.35$) and decreases at larger scale ($h \geq 0.35$).

To gain insight into this noticeable change, let us look at the samples generated by gSTAR with $N_{star} = 20$ and $N_{Cr} = 50$. Figures 5a and 5c plot the sample points of the conditional probability distribution $p(X_2|X_1)$ for $\rho = 0$ and $\rho = -0.8$, respectively. The squared markers correspond to sample points at $X_1 = 0.60$, the triangle markers correspond to sample points at $X_1 = -0.88$, and the gray circle markers

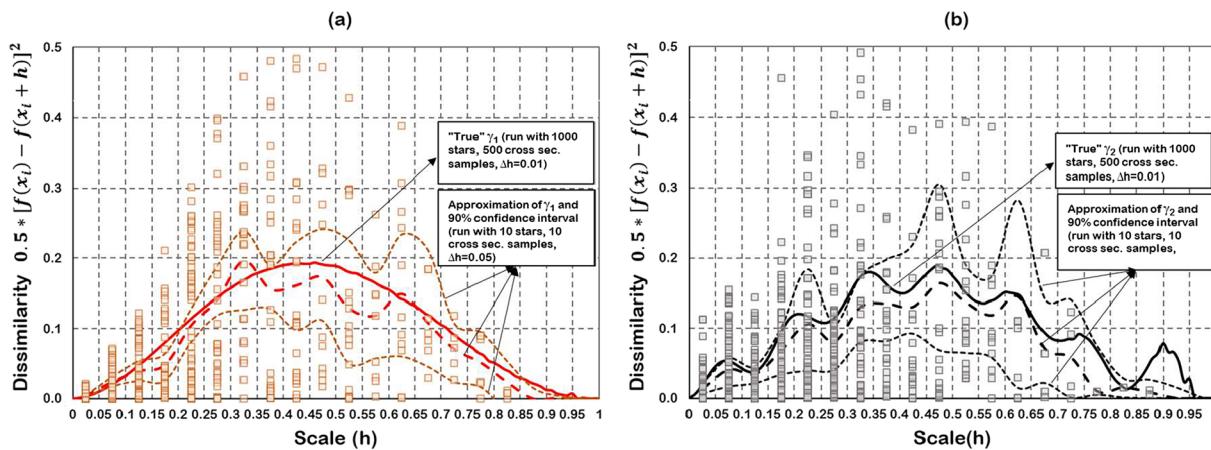


Figure 3. Computation of the directional variograms (a) for input factor 1 and (b) for input factor 2. The squared markers show the variogram clouds after clustering into different bins.

are sample points for the remaining star centers. If correlation is not considered, sample points at any cross section clearly vary at their full range $[-1; 1]$. When a linear correlation is introduced, the sample points at each cross section vary in a much narrower range, for instance, at $X_1 = 0.60$ and -0.88 , X_2 is sampled from $[-0.93; 0.16]$ and $[-0.06; 1]$, respectively. The multimodal function computed by these sample designs, therefore, has the structure and distribution shown in Figures 5b and 5d. Figures 5e and 5f provide the variogram clouds between all paired differences $\frac{1}{2}[f(X_2|X_1, X_1) - f(X_2 + h_{X_2}|X_1, X_1)]^2$ and their corresponding distances, at $X_1 = 0.6$ and $X_1 = -0.88$, respectively. The paired differences in the case of $\rho = -0.8$ are scattered in a much narrower range compared to the noncorrelation case $\rho = 0$. The estimated variogram (represented as the dashed black line in Figure 4c), which is computed by taking the average of these differences among all stars, therefore significantly decreases at larger perturbation scales due to the reduction of pairs at larger distances.

A summary of the ratio of sensitivity indices corresponding to different SA metrics (i.e., Sobol's total indices modified for correlated factor, IVARS, and VARS-TO) is shown in Figure 6. Good agreements are obtained between the VARS-TO from the proposed approach and the Sobol-TO result from the Kucherenko et al. (2012) approach for both uncorrelated and correlated cases. Specifically, factor x_1 has an equal role as factor x_2 if $\rho = 0$, while it becomes slightly less important (0.48 vs. 0.52) if $\rho = -0.8$. Introducing correlation in this problem slightly changes the results based on IVARS₁₀, IVARS₃₀, and IVARS₅₀ as well; the

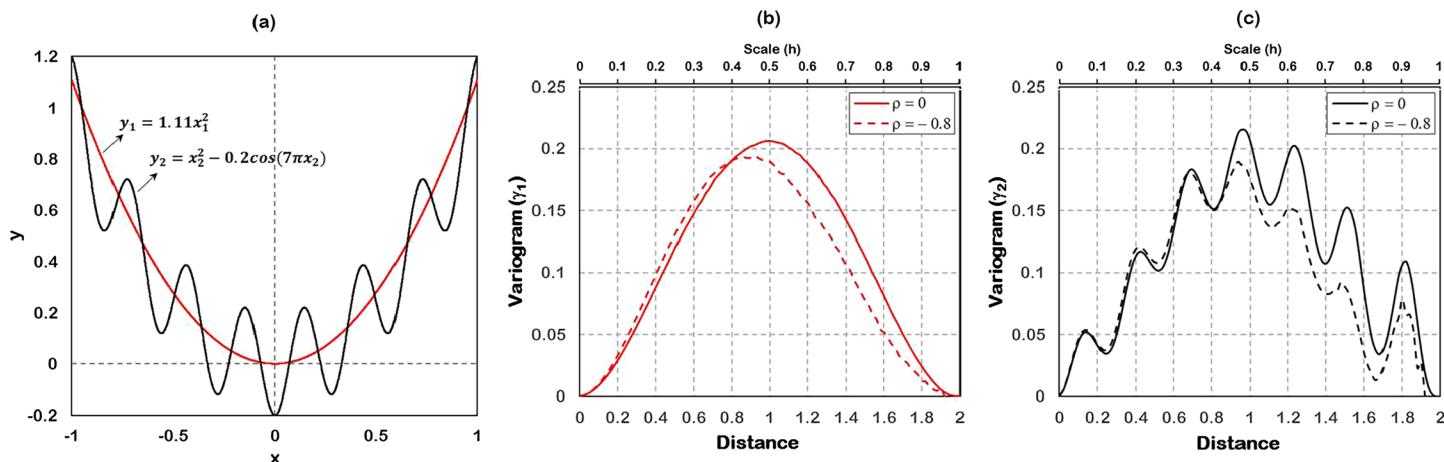


Figure 4. Performance of generalized VARS on example 1: (a) response surface structures and (b, c) the directional variograms for each factor X_1 and X_2 .

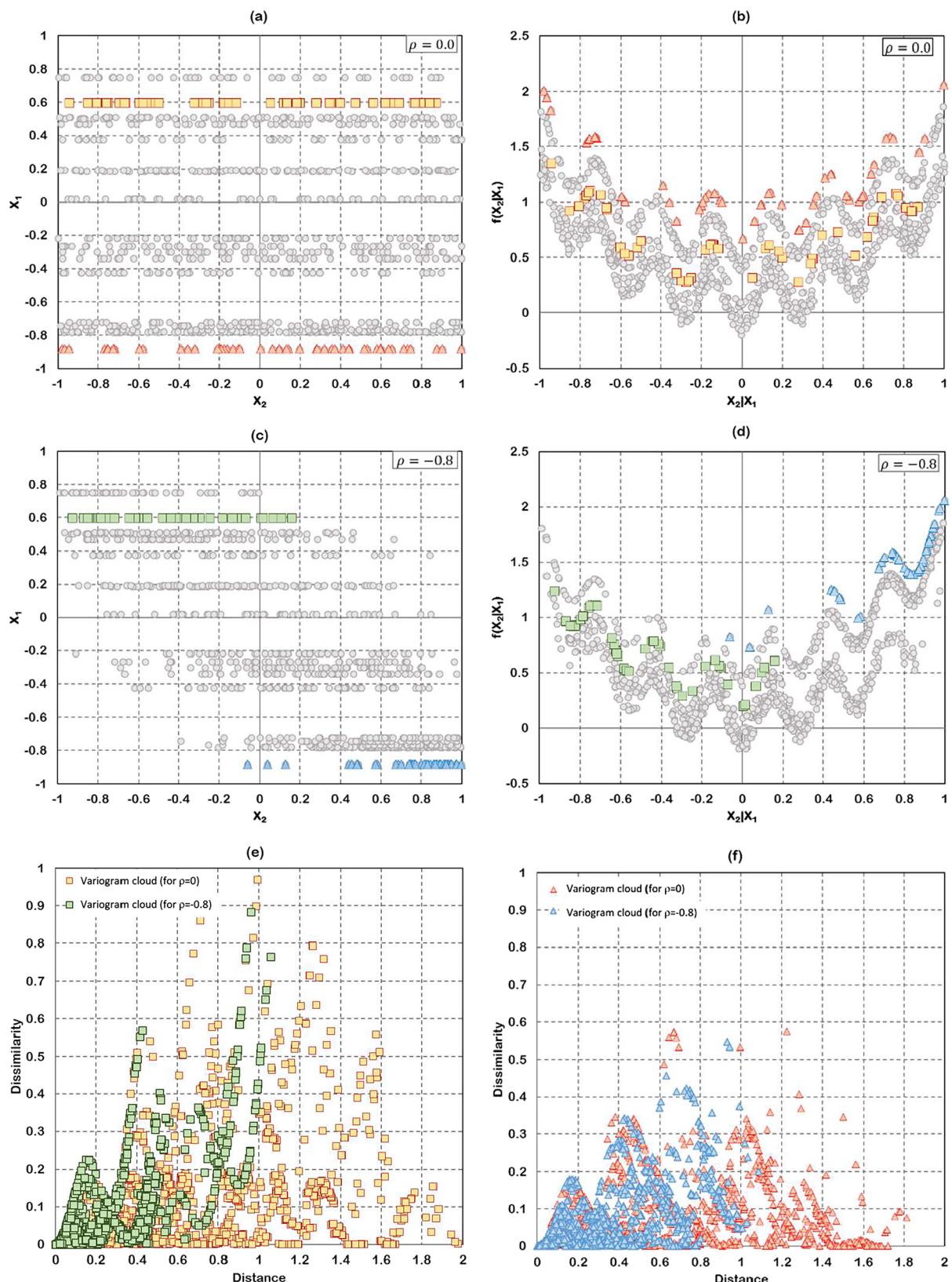


Figure 5. Analysis of correlation effects for example 2: The samples generated from $p(X_2|X_1)$ at (a) $\rho = 0$ and (c) $\rho = -0.8$, respectively; (b, d) the corresponding multimodal response surface; and the variogram cloud (paired dissimilarities) at two cross sections (e) $X_2 | (x_1 = 0.6)$ and (f) $X_2 | (x_1 = -0.88)$. Color code: Yellow and red correspond to $\rho = 0$, green and blue correspond to $\rho = -0.8$; and gray corresponds to samples at other stars.

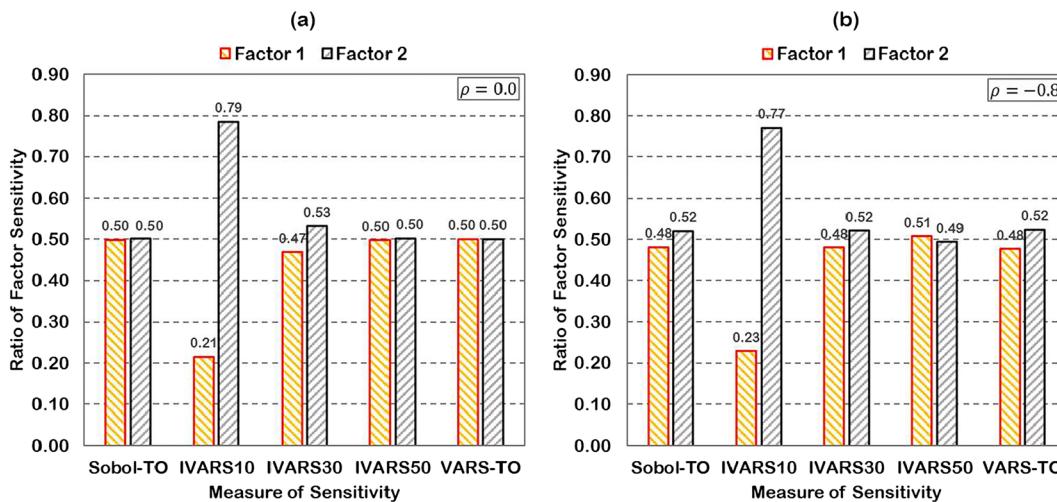


Figure 6. Summary of the ratio of sensitivity indices for a $\rho = 0$ and b $\rho = -0.8$.

importance of x_1 increases from 0.21 to 0.23 for IVARS₁₀, 0.47 to 0.48 for IVARS₃₀, and 0.50 to 0.51 for IVARS₅₀.

5.2. Example 2: $Y = \sin X_1 + 7\sin^2 X_2 + 0.1X_3^4 \sin X_1$, where $X_1, X_2, X_3 \sim U(-\pi; \pi)$

In this example, we test the performance of gSTAR-VARS for the cases of (1) uncorrelated factors and (2) correlated factors where there exists a correlation between X_1 and X_3 of $\rho_{X_1, X_3} = 0.8$. Figures 7a and 7b present the directional variograms (i.e., γ_1, γ_2 , and γ_3 , corresponding to three factors X_1, X_2 , and X_3 , respectively) of the two cases, showing correlation has a significant effect on the variograms. As X_2 is designed to be uncorrelated with X_1 and X_3 in both cases, γ_2 plotted in these figures remains unchanged. On the other hand, γ_1 moves closer to γ_3 , due to the strong correlation considered between X_1 and X_3 (Figure 7b). Figures 7c and 7d plot the integrated variograms, which are used to extract the IVARS indices. Similar to the directional variogram plot, the integrated variogram of factor X_2 stays the same in both cases. Meanwhile, when the correlation between X_1 and X_3 is considered, the integrated variogram of factor X_1 moves closer to the integrated variogram of factor X_3 , which is now entirely beneath the integrated variogram of factor X_2 , as seen in Figure 7d. Finally, Figures 7e and 7f plot the ratio of factor sensitivity based on different SA indices. These plots show a good agreement between Sobol's method and gSTAR-VARS for both the uncorrelated and correlated cases (see the ratio of sensitivity measures Sobol-TO, VARS-TO, and IVARS₅₀). Specific to the uncorrelated case, all methods provide the same importance rank for the three factors: $X_1 > X_2 > X_3$. For the correlated case, while Sobol-TO (or VARS-TO) ranks $X_2 > X_1 \approx X_3$, IVARS₅₀ ranks the factors in a slightly different way: $X_2 > X_1 > X_3$. Note that the difference between these two sensitivity indices is due to the fact that, unlike the variance-based method, the IVARS indices capture the structure of the response surface. Detailed explanation of this difference can be found in Razavi and Gupta (2015, 2016a).

6. Real Case Study

Here, we apply gSTAR-VARS to a hydrologic model in the Oldman Basin in Alberta, Canada, to identify the dominant controls of uncertainty in the estimates of design floods. The eastern slopes of the Canadian Rockies, where the Oldman Basin is located, have experienced devastating floods. For example, the costliest flood in the Canadian history occurred in this region in June 2013, with estimated losses and recovery costs of more than 6 billion CAD (Pomeroy et al., 2016). Figure 8a shows the Oldman River Basin ending at Waldron's Corner with a drainage area of 1,434.73 km². The Oldman River originates from the front ranges of the Rocky Mountains in southern Alberta and flows into the South Saskatchewan River. Historical data for this river basin are available from 1979 to 2008, with an estimated average annual precipitation of

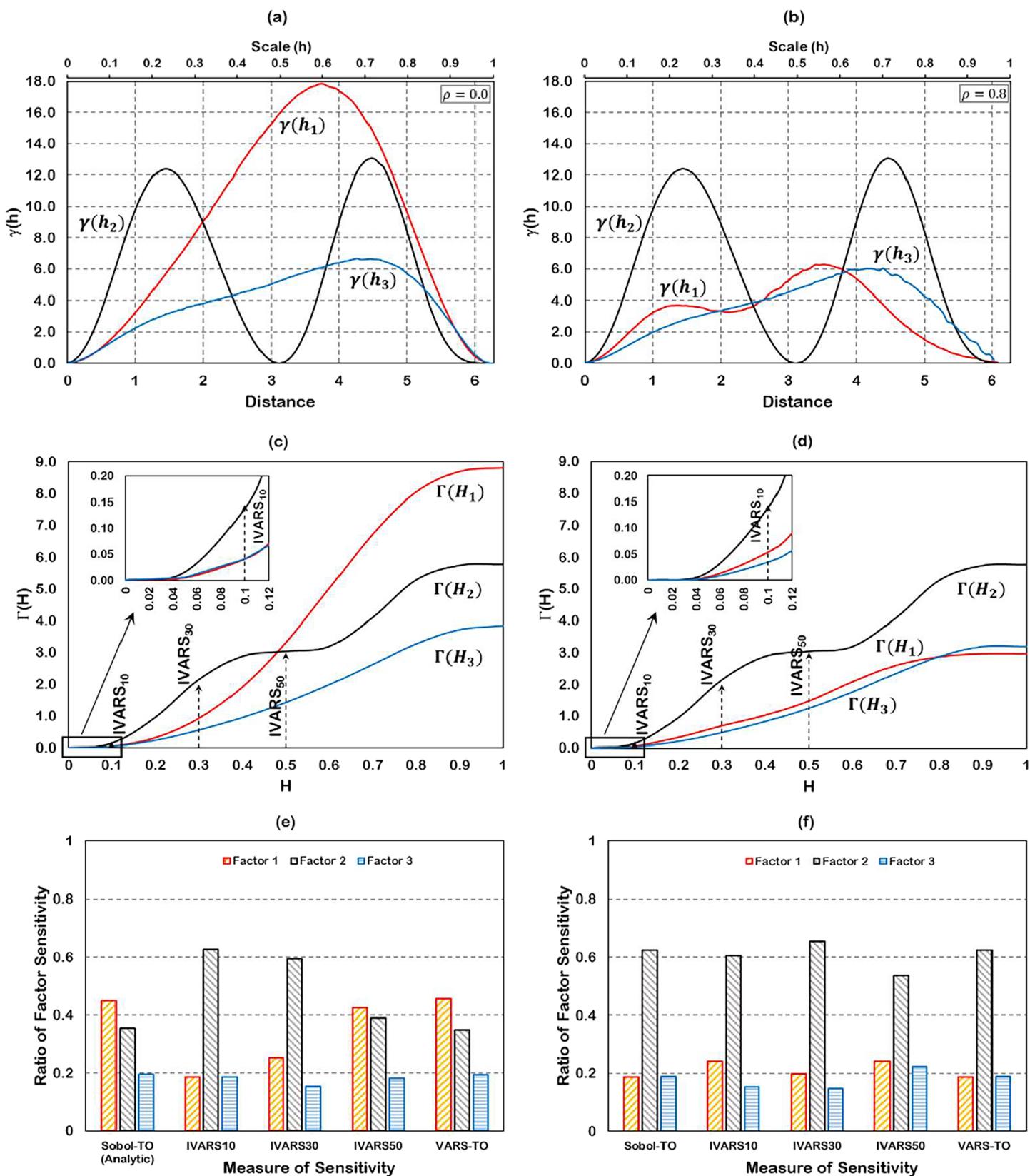


Figure 7. Performance of generalized VARS on the Ishigami function for (a, c, e) uncorrelated variables and (b, d, f) correlated variables, where (a) and (b) plot the directional variograms, (c) and (d) plot the integrated variograms, and (e) and (f) show the ratio of sensitivity factors.

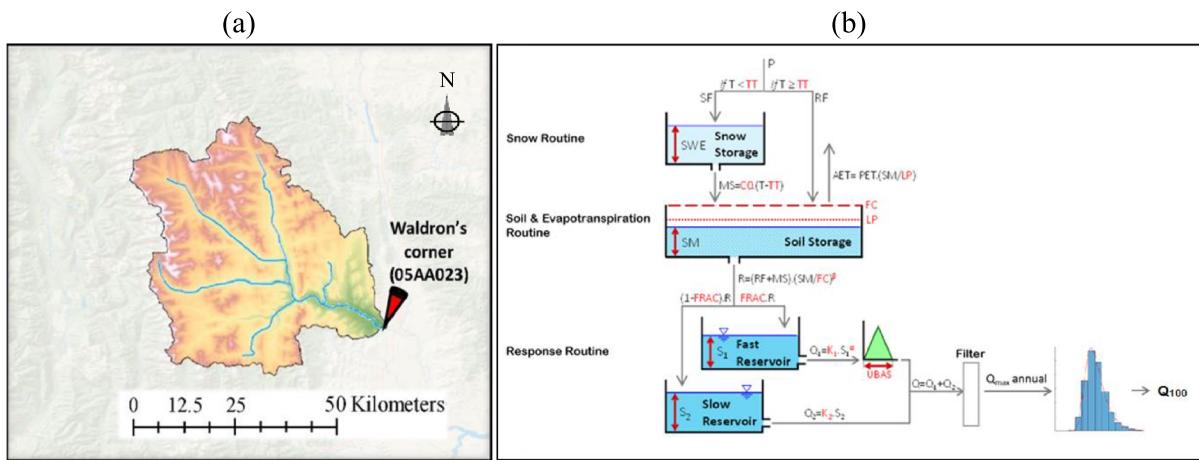


Figure 8. A Oldman River watershed and b the HBV-SASK model for flood frequency analysis.

611 mm and an annual average streamflow of $11.7 \text{ m}^3/\text{s}$ (at Gauge 05AA023). The rainfall-runoff HBV-SASK model (Razavi et al., 2019) is used to model hydrological processes in the watershed.

The HBV-SASK rainfall-runoff model is constructed from three main routines—"snow routine," "soil and evapotranspiration routine," and "response routine" as shown in Figure 8b. The snow routine represents snow accumulation and melt processes. The soil and evapotranspiration routine controls the wetness and soil moisture storage of the watershed. The response routine transforms excess water (from the soil routine) to discharge. This routine consists of two reservoirs: the fast reservoir, which controls the surface and near surface flows, and the slow reservoir, which controls the low flow in the watershed. This HBV-SASK model is coupled with a module that performs flood frequency analysis to estimate design floods with different return periods, by fitting the Pearson III distribution to annual maximum floods simulated by HBV-SASK under different parameter values. This type of distribution is suggested in Lowe and Quazi (1991) for flood frequency analysis for the Oldman River. The objective in this case study is to identify the model parameters whose uncertainties dominantly control the uncertainty in the estimates of the 100-year flood.

The parameter values of the HBV-SASK model and their associated uncertainty for this case study were estimated using the DREAM algorithm (Vrugt, 2016). DREAM is a multichain Markov chain Monte Carlo simulation algorithm that has been successfully applied in various environmental fields. This algorithm has the advantage of maintaining detailed balance and ergodicity while significantly increasing the efficiency of exploring the posterior parameter distribution, especially for complex and nonlinear models (Vrugt, 2016). A detailed description of the algorithm can be found in Vrugt (2016). In this case study, only the parameter uncertainty, represented by a multivariate distribution inferred by DREAM, was of interest. However, the procedure proposed here can be applied to consider other sources of uncertainty. The gSTAR-VARS method was applied considering the simulated 100-year flood as the model response. Details of the process of parameter estimation and the identification of dominant processes are presented in the following sections.

6.1. Parameter Estimation of the HBV-SASK Model

The HBV-SASK model is controlled by 11 parameters. Parameters C_0 (base temperature melt factor) and TT (air temperature threshold) are from the snow routine; ETF (evapotranspiration temperature-anomaly correction factor), FC (field capacity of soil), LP (multiplier to the field capacity of soil), and β (exponent in the nonlinear soil release equation) are from the soil and evapotranspiration routine; $FRAC$ (fraction of soil release entering the fast reservoir), K_1 (fast reservoir coefficient), α (exponent in the fast reservoir equation), and $UBAS$ (base of unit hydrograph for watershed routing) are components of the fast reservoir structure of the response routine; and K_2 (slow reservoir drainage rate) belongs to the slow reservoir of the response routine. More details on the HBV-SASK parameters and their initial ranges are available in Razavi et al. (2019).

Table 1
Distribution Types of Input Parameters Used in GSA

Parameter	Fitted marginal distribution (used for cases 1 and 2 in GSA)	Uniform distribution (used for case 3 in GSA)
TT	Uniform [LB = 0.94, UB = 98]	[0.94, 0.98]
C0	Normal ($\mu = 0.782$, $\sigma = 0.003$)	[0.77, 0.79]
ETF	Normal (0.126, 0.008)	[0.09, 0.16]
LP	Normal (0.67, 0.018)	[0.59, 0.73]
FC	Normal (227.53, 6.93)	[227, 280]
Beta (β)	Triangular [LB = 2.6, UB = 3.0, mode = 3.0]	[2.6, 3.0]
FRAC	Normal (0.628, 0.011)	[0.59, 0.67]
K1	Triangular [0.05, 0.054, 0.05]	[0.05, 0.054]
Alpha (α)	Normal (1.602, 0.011)	[1.56, 1.66]
K2	Normal (0.022, 0.001)	[0.017, 0.03]
UBAS	Lognormal (0.05, 0.04)	[1.0, 1.26]

The multivariate distribution of parameter values of the HBV-SASK model was estimated using the DREAM algorithm on the 15-year data from January 1982 to December 1997 and validated on data of the remaining years—for brevity, details of calibration and validation efforts including hydrographs are not reported here. For this case study, we ran the DREAM algorithm with 11 different chains and 15,000 generations while keeping other DREAM parameters at their default values. We used the parameter bounds listed in Table S1 (supporting information) as a priori conditions and applied the Gaussian likelihood function (equation (10)) proposed by Vrugt (2016):

$$\mathcal{L}(X|Q_{obs}) = -\frac{n}{2} \log \left\{ \sum_{t=1}^n e_t(X|Q_{obs})^2 \right\}, \quad (10)$$

where $X = \{X_1, X_2, \dots, X_{11}\}$ is the vector of the parameters, Q_{obs} is the vector of observed streamflow, $e_t(X)$ is the residual at time t , and n is the length of the simulation period.

The histograms of the parameters obtained a posteriori are shown in Figure S5 (supporting information). These histograms were obtained using the last 6,000 samples, after meeting the convergence criterion (i.e., the multichain \hat{R} -statistic) of sampled chains. This resulted in a set of 66,000 sample points for each parameter. It is seen that the parameter ranges are significantly reduced from their a priori ranges, for example, from [-4,4] to [0.94,0.98] for TT and from [50,500] to [227,280] for FC. For use with GSA, we fitted theoretical distributions to the histograms, as summarized in Table 1. The fitted distribution for each parameter is selected from six theoretical distribution types (uniform, lognormal, normal, triangular, exponential, and generalized extreme value distributions) and based on the chi-square test for the goodness of fit. The fitted distributions also need to ensure that the Nataf transformation can be implemented. Further, we extracted the correlation structure among the parameters by the linear correlation matrix reported in Table 2. Some relatively strong pairwise relationships exist between several parameters. The highest correlation (0.71) is inferred between β and LP, while other relatively strong correlations can be seen among FC and LP, TT and C0, and α and FRAC.

6.2. SA Setup and Results

To demonstrate the applicability of the gSTAR-VARS method and evaluate the effects of the correlation matrix and the distribution types on the sensitivity results, we ran gSTAR-VARS for three different cases: Case 1—the model parameters are correlated and have their nonuniform distributions as shown in Table 1 (column 2); Case 2—the correlation is removed, and only the nonuniform distributional characteristics of the parameters are considered; and Case 3—all parameters were assumed to be uniformly

Table 2
The Correlation Structure Between Calibrated Parameters of the HBV-SASK Model Applied to the Oldman River Basin

Parameter	TT	C0	ETF	LP	FC	β	FRAC	K1	α	K2	UBAS
TT	1	0.65	0	0	0	0	0	0	0	0	0
C0		1	0	0	0	0	0	0	0	0.12	0
ETF			1	0.12	-0.18	0.13	0	0	0	-0.22	0
LP				1	0.54	0.71	-0.14	0	0	0	0
FC					1	0.34	0.2	0.11	0	0.38	0
β						1	-0.11	0	0	-0.13	0
FRAC							1	0	-0.69	-0.39	-0.19
K1								1	-0.34	0	0
α									1	0.41	0.4
K2										1	0.14
UBAS											1

Note: Weak correlations ($|\rho_{i,j}| < 0.1$) are negligible and are replaced by zero in this table.

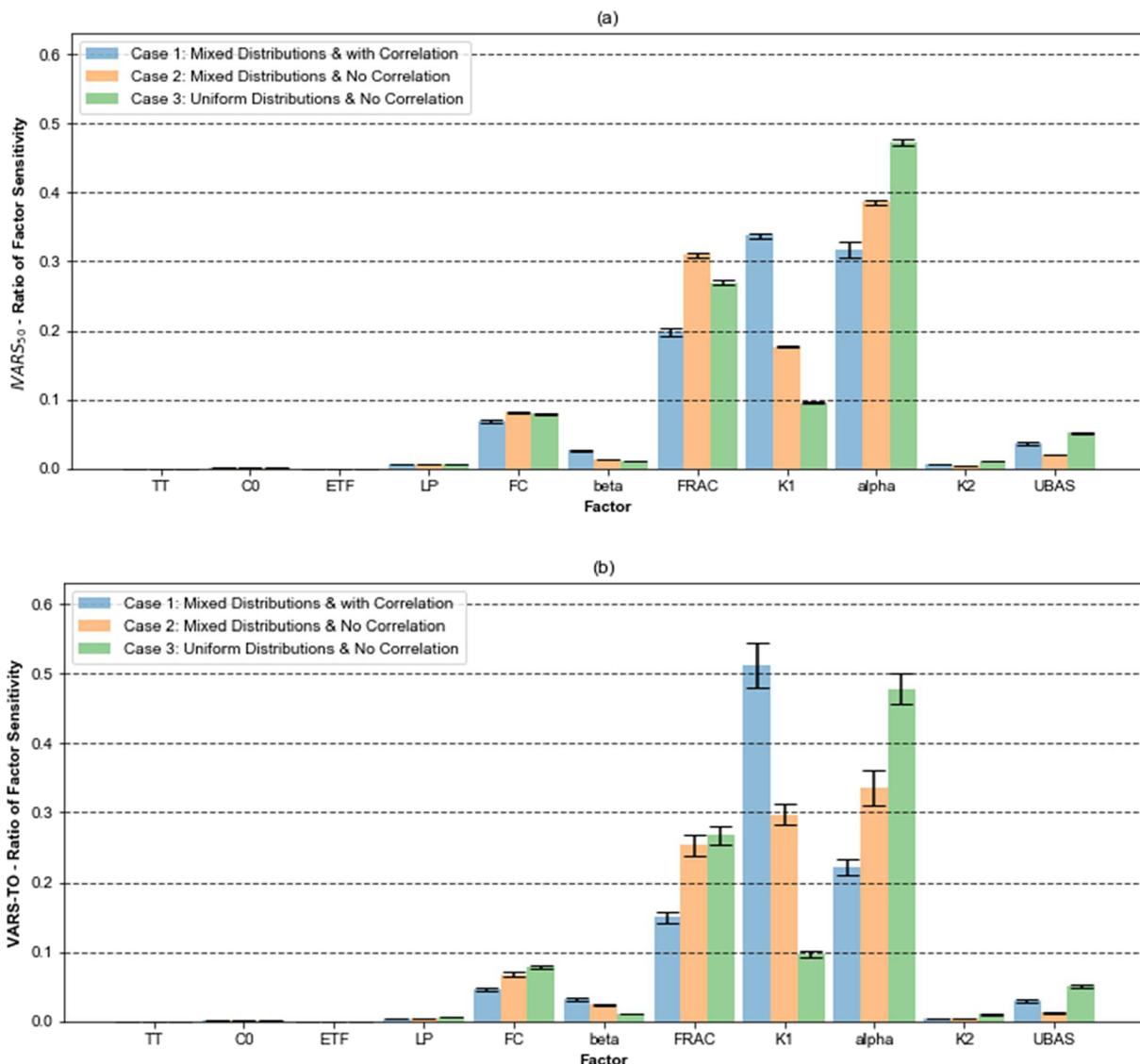


Figure 9. Ratio of sensitivity factor provided by (a) IVARS₅₀ and (b) VARS-TO for case 1 (parameters are mixed distributed and correlated), case 2 (parameters are mixed distributed and independent), and case 3 (parameters are uniformly distributed and independent).

distributed and independent; the upper and lower bounds of the parameters are the maximum and minimum values of the posterior sample obtained from the parameter estimation process. Case 3 is constructed to emulate the original implementation of VARS, called STAR-VARS (Razavi & Gupta, 2016b).

The proposed gSTAR-VARS method was run on the HBV-SASK model with different settings to evaluate its convergence: the number of stars, N_{Star} , was set to 100, 200, 500, and 1,000; the number cross-sectional points, N_{cr} , was set to 10, 20, and 50; the number of parameters $N_{fac} = 11$; and the resolution $\Delta h = 0.05$ to compute the variograms and covariograms. It is seen that when the number of cross-sectional sample points reaches 50, and the number of star centers is larger than 500, the method starts to converge to a unique sensitivity result (Figure S6 in the supporting information). Therefore, we used $N_{Star} = 1,000$ and $N_{cr} = 50$ to evaluate this case study. The results of Case 1 are discussed below, while the comparisons of the three cases and the lessons learned are presented in the next section.

Figure 9a provides the IVARS₅₀ indices and their respective 90% confidence intervals, shown as error bars, using bootstrap within each case. An alternative to bootstrap could be the Model Variable Augmentation method developed by Mai and Tolson (2019) (not performed here). Results of Case 1 show that the flood

estimates are most sensitive to uncertainty in parameters that govern the fast reservoir structure of the HBV-SASK model. Specifically, the fast reservoir coefficient (K_1) is the most important parameter, contributing a ratio of 0.337 (i.e., 33.7%) to the sum of the sensitivity indices. The exponent of the fast reservoir equation (α) is ranked as the second most important parameter (31.6%), followed by the fraction of soil release entering the fast reservoir (FRAC) (19.8%). These results are consistent with the role of these parameters, which is built to control the surface and shallow subsurface flows.

The importance ranking of “moderately important” parameters (i.e., the parameters having contributions of roughly between 1% and 10% to the model output variability) according to the ratios of factor sensitivity is as follows: FC (5.6%) > UBAS (3.6%) > β (2.6%). Except UBAS, which governs watershed routing as a part of the fast reservoir structure, FC and β belong to the soil routine of the HBV-SASK model. These parameters determine the conversion of net precipitation into soil moisture recharge and runoff. The results indicate that, in the soil process, the field capacity of soil (FC; the maximum amount of water that the soil can retain) is more influential on the flood than is the shape parameter of the soil release equation (β).

The remaining factors have the least impact (<1%) on the 100-year flood estimate, with the ratios of factor sensitivity as follows: K_2 (0.67%) > LP (0.56%) > C_0 (0.2%) > TT (0.03%) > ETF (0.012%). Parameter K_2 belongs to the response routine but only controls the low flows of the watershed via the slow reservoir, parameters LP and ETF determine the limit of daily potential evapotranspiration in relation to the available moisture in the evapotranspiration routine, and parameters TT and C_0 define the snow processes in the snow routine. Therefore, these parameters only have an indirect, typically limited impact on the discharge at peak times.

Figure 9b provides the sensitivity results of the three cases in terms of the VARS-TO indices. For Case 1, while the ratios of factor sensitivity are different from those of IVARS₅₀, the importance rankings based on the two indices are consistent. The only change in rankings is around parameters β and UBAS, where β (3.1%) is ranked to be more important than UBAS (2.9%), according to VARS-TO. However, the 90% confidence intervals show that the ranks of these two parameters are almost interchangeable, as the contribution of β ranges from 3% to 3.3%, while the contribution of UBAS ranges from 2.7% to 3.1%.

6.3. Effects of Parameter Correlation and Distribution Type on Sensitivity Results

The effect of correlation on SA can be seen by comparing the results from Case 1 (correlation is considered) and Case 2 (correlation is not considered), as shown in Figure 9. The correlation between the parameters affects not only the magnitude of the relative sensitivity but also the importance rank of the parameters. For example, according to IVARS₅₀, while K_1 (33.7%) is ranked the most important parameter in Case 1, it becomes the third most important parameter in Case 2 (17.6%). Similarly, parameter α , which is ranked the second most important in Case 1, becomes the most important in Case 2. According to VARS-TO, K_1 and α are distinctively ranked the first and second in Case 1 (with contributions of $51 \pm 3\%$ and $22 \pm 2\%$, respectively), while their ranks in Case 2 are switched.

Finally, the impact of the distribution type on the sensitivity results is evident by comparing the results of Case 2 and Case 3. For example, the relative importance of parameter α under the uniform distribution assumption (Case 3) is significantly higher than that under the normal distribution assumption (Case 2). Similarly, UBAS, assigned a lognormal distribution in Case 2, becomes more important in Case 3 where a uniform distribution is assigned.

This case study shows that the impact of correlations and nonuniformity of the parameters can be large. The extent of such impacts, in general, depends on the form, nonlinearity, and nonadditivity (interactions) of the response surface, the magnitude of correlations, and distribution types.

7. Robustness of gSTAR-VARS

In this section, we assess the “robustness” of the proposed method against the method of Kucherenko et al. (2012) on the Ishigami function and the HBV-SASK model. Robustness here is defined as the insensitivity of the results to sampling variability (e.g., because of change in seed numbers in random number generators used for sampling). To do so, we used 3,750 sample points for the Ishigami function and 23,000 sample points for the HBV-SASK case study, arbitrarily. To this end, we ran each method 100 times using different sampled

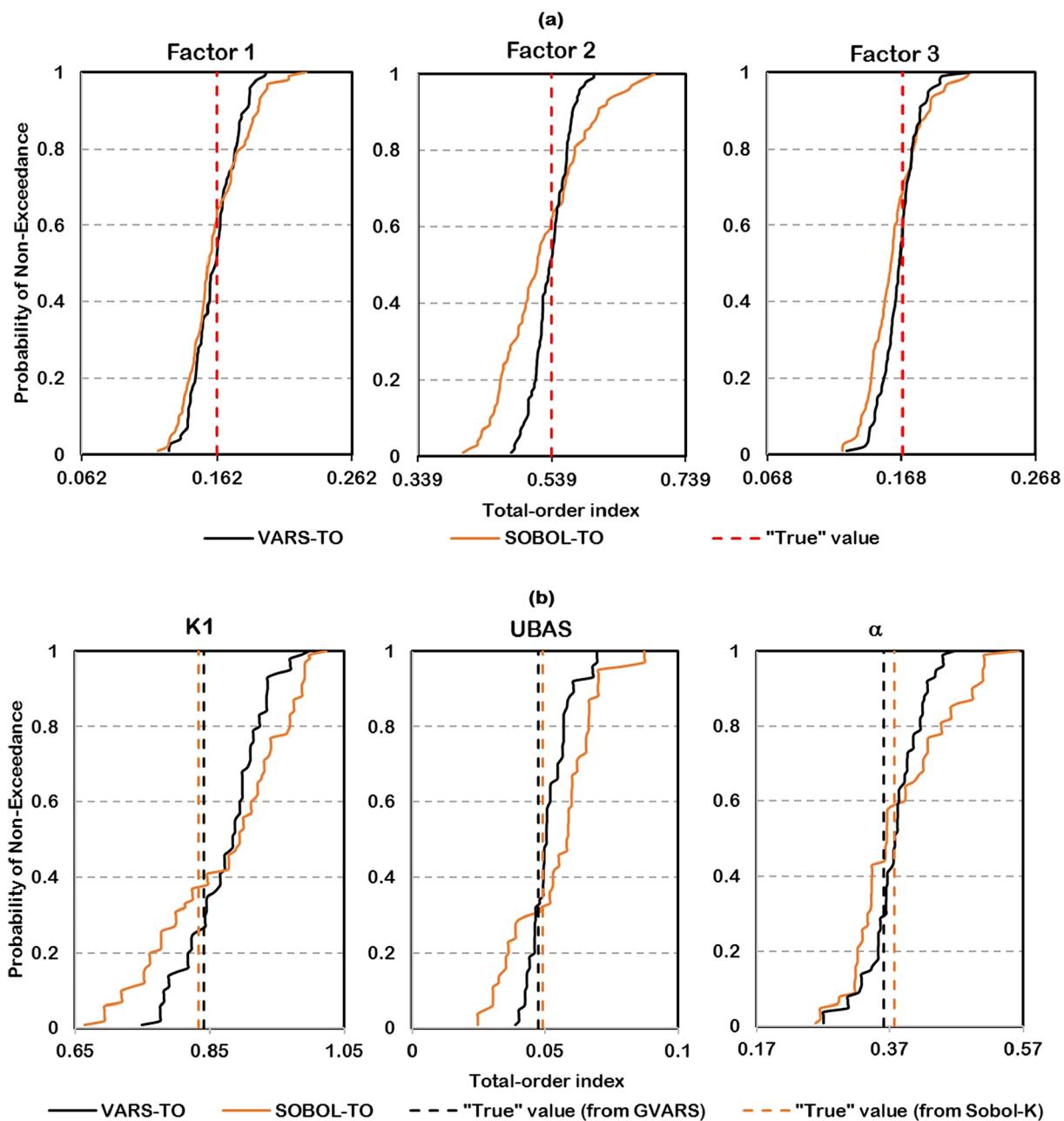


Figure 10. An assessment of the robustness of gSTAR-VARS compared with the Sobol method. Each subplot shows the cumulative distribution functions over 100 independent trials (with different initial random seeds) of the total-order index for different factors. (a) the evaluation for example 2, in which a low computational budget of 3,750 model runs is implemented. (b) the evaluation for the case study; a computational budget of 23,000 model runs is used.

points (using different random seeds for sampling). Figure 10 shows the empirical CDFs of the 100 independent estimates of VARS-TO and Sobol-TO; the top row shows the parameters of the Ishigami function, while the bottom row shows three select parameters of the HBV-SASK model. The results show that gSTAR-VARS is more stable and robust than the alternative method within the same sample size (computational budget), as the CDFs obtained by gSTAR-VARS demonstrates less variability (narrower range) while relatively symmetrically covering the “true” values of the sensitivity index.

8. Concluding Remarks

This paper proposed a novel approach for the GSA of models with correlated, nonuniformly distributed factors. The proposed approach built on the theory of VARS (Razavi & Gupta, 2016a, 2016b) and extended that

to the case of multivariate distributions. For numerical applications, we introduced a generalized star sampling strategy (gSTAR) that applied the inverse Nataf transformation to sample the joint and conditional distributions of factors. gSTAR-VARS was able to generate different sensitivity indices, including the IVARS indices such as total-variogram effects, variance-based total-order effects (VARS-TO), and derivative-based elementary effects (results not shown here) for a range of distribution types and correlation structures.

The applicability of the gSTAR-VARS method was first demonstrated using two test functions. In these examples, we achieved a very good agreement between the VARS-TO index obtained from gSTAR-VARS and that of the modified Sobol's method from Kucherenko et al. (2012), which accounts for correlated factors. In a third example (shown in the supporting information), we obtained a perfect match with the analytical result. In addition, comparison of the directional variograms with and without considering correlation effects in these examples provided a comprehensive insight into the role of correlation effects in interpreting the response surface structure. Such insights, which are conveyed through the form of directional variograms, can currently only be explored via the VARS framework.

The proposed gSTAR-VARS method was also evaluated in a case study with the HBV-SASK hydrologic model, aimed at identifying dominant sources of parameter uncertainty on the estimates of the 100-year flood in the Oldman River Basin (Canada). In this case study, we first implemented the DREAM algorithm (Vrugt, 2016) to infer the multivariate distribution of model parameters based on the historical observations. The results showed that the gSTAR-VARS method is able to characterize how uncertainty in the different model parameters controls the uncertainty in model response. Furthermore, additional evaluations showed the significance of accounting for correlations and appropriate parameter distribution types on the results of GSA.

Further research is needed to test the proposed method and its efficiency and robustness across other types of problems. Furthermore, new approaches and methods that can handle parameters with any empirical distribution are needed as the current methods are limited to analytical distributions. As always, we invite discussion and collaboration with others interested in SA, model development, model diagnostics, and related issues.

Acknowledgments

The authors gratefully acknowledge the supports from the Global Water Futures (GWF)-Integrated Modeling Program for Canada (IMPC), funded by the Canada First Research Excellence Fund, and from Razavi's NSERC Discovery Grant. We wish to thank Sergei Kucherenko, Stefano Tarantola, and Paola Annoni for providing the code of the modified Sobol's method. We also thank the three anonymous reviewers and appreciate their valuable comments that help us significantly improve the manuscript. The source code of the gSTAR-VARS method is available at <https://vars-tool.com> or at the online repository Figshare (at https://adelaide.figshare.com/articles/Variogram_method_for_Global_Sensitivity_Analysis/10007465).

References

- Ahn, H. (1996). Sensitivity for correlated input variables and propagated errors in evapotranspiration estimates from a humid region. *Water Resources Research*, 32(8), 2507–2516. <https://doi.org/10.1029/96WR01498>
- Chu-Agor, M. L., Muñoz-Carpena, R., Kiker, G., Emanuelsson, A., & Linkov, I. (2011). Exploring vulnerability of coastal habitats to sea level rise through global sensitivity and uncertainty analyses. *Environmental Modelling & Software*, 26(5), 593–604. <https://doi.org/10.1016/j.envsoft.2010.12.003>
- Chu, Y., & Hahn, J. (2007). Parameter set selection for estimation of nonlinear dynamic systems. *AIChE Journal*, 53, 2858–2870. <https://doi.org/10.1002/aic.11295>
- Cosby, B. J., Hornberger, G. M., Clapp, R. B., & Ginn, T. R. (1984). A statistical exploration of the relationships of soil moisture characteristics to the physical properties of soils. *Water Resources Research*, 20(6), 682–690. <https://doi.org/10.1029/WR020i006p00682>
- Da Veiga, S., Wahl, F., & Gamboa, F. (2009). Local polynomial estimation for sensitivity analysis on models with correlated inputs. *Technometrics*, 51(4), 452–463. <https://doi.org/10.1198/TECH.2009.08124>
- Do, N. C., Simpson, A. R., Deuerlein, J. W., & Piller, O. (2016). Calibration of water demand multipliers in water distribution systems using genetic algorithms. *Journal of Water Resources Planning and Management*, 142(11), 04016044. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000691](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000691)
- Do, N. C., Simpson, A. R., Deuerlein, J. W., & Piller, O. (2018). Locating inadvertently partially closed valves in water distribution systems. *Journal of Water Resources Planning and Management*, 144(8), 04018039. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000958](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000958)
- Gupta, H. V., & Razavi, S. (2018). Revisiting the basis of sensitivity analysis for dynamical earth system models. *Water Resources Research*, 54, 8692–8717. <https://doi.org/10.1029/2018WR022668>
- Gupta, H. V., Wagener, T., & Liu, Y. (2008). Reconciling theory with observations: Elements of a diagnostic approach to model evaluation. *Hydrological Processes*, 22(18), 3802–3813. <https://doi.org/10.1002/hyp.6989>
- Haghnegahdar, A., Razavi, S., Yassin, F., & Wheater, H. (2017). Multicriteria sensitivity analysis as a diagnostic tool for understanding model behaviour and characterizing model uncertainty. *Hydrological Processes*, 31(25), 4462–4476. <https://doi.org/10.1002/hyp.11358>
- Homma, T., & Saltelli, A. (1996). Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety*, 52(1), 1–17. [https://doi.org/10.1016/0951-8320\(96\)00002-6](https://doi.org/10.1016/0951-8320(96)00002-6)
- Kucherenko, S., Tarantola, S., & Annoni, P. (2012). Estimation of global sensitivity indices for models with dependent variables. *Computer Physics Communications*, 183(4), 937–946. <https://doi.org/10.1016/j.cpc.2011.12.020>
- Lemke, L. D., Abriola, L. M., & Goovaerts, P. (2004). Dense nonaqueous phase liquid (DNAPL) source zone characterization: Influence of hydraulic property correlation on predictions of DNAPL infiltration and entrapment. *Water Resources Research*, 40, W01511. <https://doi.org/10.1029/2003WR001980>
- Li, G., Rabitz, H., Yelvington, P. E., Oluwole, O. O., Bacon, F., Kolb, C. E., & Schoendorf, J. (2010). Global sensitivity analysis for systems with independent and/or correlated inputs. *The Journal of Physical Chemistry A*, 114(19), 6022–6032. <https://doi.org/10.1021/jp9096919>

- Lidén, R., & Harlin, J. (2000). Analysis of conceptual rainfall-runoff modelling performance in different climates. *Journal of Hydrology*, 238(3), 231–247. [https://doi.org/10.1016/S0022-1694\(00\)00330-9](https://doi.org/10.1016/S0022-1694(00)00330-9)
- Liu, P. L., & Der Kiureghian, A. (1986). Multivariate distribution models with prescribed marginals and covariances. *Probabilistic Engineering Mechanics*, 1(2), 105–112. [https://doi.org/10.1016/0266-8920\(86\)90033-0](https://doi.org/10.1016/0266-8920(86)90033-0)
- Lowe, S., & Quazi, M. (1991). Oldman River at Fort Macleod Flood Risk Mapping Study, Alberta flood damage reduction program, <https://open.alberta.ca/publications/fort-macleod-oldman-river-flood-hazard-study>
- Mai, J., & Tolson, B. A. (2019). Model variable augmentation (MVA) for diagnostic assessment of sensitivity analysis results. *Water Resources Research*, 55, 2631–2651. <https://doi.org/10.1029/2018WR023382>
- Manache, G., & Melching, C. S. (2008). Identification of reliable regression- and correlation-based sensitivity measures for importance ranking of water-quality model parameters. *Environmental Modelling & Software*, 23(5), 549–562. <https://doi.org/10.1016/j.envsoft.2007.08.001>
- Mara, T. A., & Tarantola, S. (2012). Variance-based sensitivity indices for models with dependent inputs. *Reliability Engineering & System Safety*, 107, 115–121. <https://doi.org/10.1016/j.ress.2011.08.008>
- Mara, T. A., Tarantola, S., & Annoni, P. (2015). Non-parametric methods for global sensitivity analysis of model output with dependent inputs. *Environmental Modelling & Software*, 72, 173–183. <https://doi.org/10.1016/j.envsoft.2015.07.010>
- Marchant, B. P., & Lark, R. M. J. M. G. (2004). Estimating variogram uncertainty. *Mathematical Geology*, 36(8), 867–898. <https://doi.org/10.1023/B:MATG.0000048797.08986.a7>
- Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2), 161–174. <https://doi.org/10.2307/1269043>
- Muleta, M. K., & Nicklow, J. W. (2005). Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model. *Journal of Hydrology*, 306(1), 127–145. <https://doi.org/10.1016/j.jhydrol.2004.09.005>
- Nossent, J., Elsen, P., & Bauwens, W. (2011). Sobol's sensitivity analysis of a complex environmental model. *Environmental Modelling & Software*, 26(12), 1515–1525. <https://doi.org/10.1016/j.envsoft.2011.08.010>
- Pan, F., Zhu, J., Ye, M., Pachepsky, Y. A., & Wu, Y.-S. (2011). Sensitivity analysis of unsaturated flow and contaminant transport with correlated parameters. *Journal of Hydrology*, 397(3), 238–249. <https://doi.org/10.1016/j.jhydrol.2010.11.045>
- Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., & Wagener, T. (2016). Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modelling & Software*, 79, 214–232. <https://doi.org/10.1016/j.envsoft.2016.02.008>
- Pianosi, F., & Wagener, T. (2018). Distribution-based sensitivity analysis from a generic input-output sample. *Environmental Modelling & Software*, 108, 197–207. <https://doi.org/10.1016/j.envsoft.2018.07.019>
- Plischke, E., Borgonovo, E., & Smith, C. L. (2013). Global sensitivity measures from given data. *European Journal of Operational Research*, 226(3), 536–550. <https://doi.org/10.1016/j.ejor.2012.11.047>
- Pomeroy, J. W., Stewart, R. E., & Whitfield, P. H. (2016). The 2013 flood event in the South Saskatchewan and Elk River basins: Causes, assessment and damages. *Canadian Water Resources Journal/Revue canadienne des ressources hydriques*, 41(1–2), 105–117. <https://doi.org/10.1080/07011784.2015.1089190>
- Rakovec, O., Hill, M. C., Clark, M. P., Weerts, A. H., Teuling, A. J., & Uijlenhoet, R. (2014). Distributed evaluation of local sensitivity analysis (DELSA), with application to hydrologic models. *Water Resources Research*, 50, 409–426. <https://doi.org/10.1002/2013WR014063>
- Razavi, S., & Gupta, H. V. (2015). What do we mean by sensitivity analysis? The need for comprehensive characterization of “global” sensitivity in earth and environmental systems models. *Water Resources Research*, 51, 3070–3092. <https://doi.org/10.1002/2014WR016527>
- Razavi, S., & Gupta, H. V. (2016a). A new framework for comprehensive, robust, and efficient global sensitivity analysis: 1. Theory. *Water Resources Research*, 52, 423–439. <https://doi.org/10.1002/2015WR017558>
- Razavi, S., & Gupta, H. V. (2016b). A new framework for comprehensive, robust, and efficient global sensitivity analysis: 2. Application. *Water Resources Research*, 52, 440–455. <https://doi.org/10.1002/2015WR017559>
- Razavi, S., Sheikholeslami, R., Gupta, H. V., & Haghnegahdar, A. (2019). VARS-TOOL: A toolbox for comprehensive, efficient, and robust sensitivity and uncertainty analysis. *Environmental Modelling & Software*, 112, 95–107. <https://doi.org/10.1016/j.envsoft.2018.10.005>
- Razavi, S., Tolson, B. A., & Burn, D. H. (2012). Numerical assessment of metamodelling strategies in computationally intensive optimization. *Environmental Modelling & Software*, 34, 67–86. <https://doi.org/10.1016/j.envsoft.2011.09.010>
- Saltelli, A., & Annoni, P. (2010). How to avoid a perfunctory sensitivity analysis. *Environmental Modelling & Software*, 25(12), 1508–1517. <https://doi.org/10.1016/j.envsoft.2010.04.012>
- Saltelli, A., Ratto, M., Tarantola, S., & Campolongo, F. (2006). Sensitivity analysis practices: Strategies for model-based inference. *Reliability Engineering & System Safety*, 91(10), 1109–1125. <https://doi.org/10.1016/j.ress.2005.11.014>
- Saltelli, A., Tarantola, S., Campolongo, F., & Ratto, M. (2004). Sensitivity analysis in practice: A guide to assessing scientific models, Halsted Press. <https://doi.org/10.1002/0470870958>
- SAS Institute Inc (1999). *SAS/STAT User's Guide*. Cary, NC, USA: SAS Institute Inc. <http://www.math.wpi.edu/saspdf/stat/chap70.pdf>
- Sheikholeslami, R., Razavi, S., Gupta, H. V., Becker, W., & Haghnegahdar, A. (2019). Global sensitivity analysis for high-dimensional problems: How to objectively group factors and measure robustness and convergence while reducing computational cost. *Environmental Modelling & Software*, 111, 282–299. <https://doi.org/10.1016/j.envsoft.2018.09.002>
- Smith, T. E. (2016). Notebook on Spatial Data Analysis [online] <http://www.seas.upenn.edu/~ese502/#notebook>
- Sobol, I. M. (1993). Sensitivity estimates for nonlinear mathematical models. *MMCE*, 4(1), 407–414. 1061-7590/93/04407-008
- Song, X., Zhang, J., Zhan, C., Xuan, Y., Ye, M., & Xu, C. (2015). Global sensitivity analysis in hydrological modeling: Review of concepts, methods, theoretical framework, and applications. *Journal of Hydrology*, 523, 739–757. <https://doi.org/10.1016/j.jhydrol.2015.02.013>
- Spear, R. C., & Hornberger, G. M. (1980). Eutrophication in peat inlet—II. Identification of critical uncertainties via generalized sensitivity analysis. *Water Research*, 14(1), 43–49. [https://doi.org/10.1016/0043-1354\(80\)90040-8](https://doi.org/10.1016/0043-1354(80)90040-8)
- Tarantola, S., & Mara, T. A. (2017). Variance-based sensitivity indices of computer models with dependent inputs: The Fourier amplitude sensitivity test. *International Journal for Uncertainty Quantification*, 7(6), 511–523. <https://doi.org/10.1615/Int.J.UncertaintyQuantification.2017020291>
- van Griensven, A., Meixner, T., Grunwald, S., Bishop, T., Diluzio, M., & Srinivasan, R. (2006). A global sensitivity analysis tool for the parameters of multi-variable catchment models. *Journal of Hydrology*, 324(1), 10–23. <https://doi.org/10.1016/j.jhydrol.2005.09.008>

- Vrugt, J. A. (2016). Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environmental Modelling & Software*, 75, 273–316. <https://doi.org/10.1016/j.envsoft.2015.08.013>
- Wagener, T., Boyle, D. P., Lees, M. J., Wheater, H. S., Gupta, H. V., & Sorooshian, S. (2001). A framework for development and application of hydrological models. *Hydrology and Earth System Sciences*, 5(1), 13–26. <https://doi.org/10.5194/hess-5-13-2001>
- Xu, C., & Gertner, G. Z. (2008). Uncertainty and sensitivity analysis for models with correlated parameters. *Reliability Engineering & System Safety*, 93(10), 1563–1573. <https://doi.org/10.1016/j.ress.2007.06.003>
- Zhu, J. (2012). Sensitivity of advective travel time of contaminants to correlated formation porosities. *Hydrogeology Journal*, 20(1), 135–141. <https://doi.org/10.1007/s10040-011-0788-0>