

UC Irvine

UC Irvine Previously Published Works

Title

Probabilistic Sensitivity Analysis With Dependent Variables: Covariance-Based Decomposition of Hydrologic Models

Permalink

<https://escholarship.org/uc/item/7n69665p>

Journal

Water Resources Research, 59(4)

ISSN

0043-1397

Authors

Gao, Yifu
Sahin, Abdullah
Vrugt, Jasper A

Publication Date

2023-04-01

DOI

10.1029/2022wr032834

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Water Resources Research[®]



RESEARCH ARTICLE

10.1029/2022WR032834

Probabilistic Sensitivity Analysis With Dependent Variables: Covariance-Based Decomposition of Hydrologic Models

Yifu Gao¹ , Abdullah Sahin¹, and Jasper A. Vrugt¹ 

¹Department of Civil and Environmental Engineering, University of California, Irvine, CA, USA

Key Points:

- We use high-dimensional model representation for variance-based sensitivity analysis with dependent input variables
- We use this method to disentangle the parameters' structural and correlative contributions across the posterior distribution
- Correlation can profoundly affect input variable importance, optimal experimental design and physical underpinning of sensitivity indices

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

J. A. Vrugt,
jasper@uci.edu

Citation:

Gao, Y., Sahin, A., & Vrugt, J. A. (2023). Probabilistic sensitivity analysis with dependent variables: Covariance-based decomposition of hydrologic models. *Water Resources Research*, 59, e2022WR032834. <https://doi.org/10.1029/2022WR032834>

Received 17 MAY 2022

Accepted 21 MAR 2023

Author Contributions:

Conceptualization: Jasper A. Vrugt
Data curation: Yifu Gao, Abdullah Sahin, Jasper A. Vrugt
Formal analysis: Yifu Gao
Investigation: Yifu Gao, Abdullah Sahin, Jasper A. Vrugt
Methodology: Yifu Gao, Abdullah Sahin, Jasper A. Vrugt
Resources: Jasper A. Vrugt
Software: Abdullah Sahin, Jasper A. Vrugt
Supervision: Jasper A. Vrugt
Validation: Yifu Gao, Abdullah Sahin, Jasper A. Vrugt
Visualization: Yifu Gao

© 2023. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Abstract Variance-based analysis has emerged as method of choice for quantifying the sensitivity of the output, y , of a scalar-valued square-integrable function, $f \in L^2(\mathcal{X})$, to its $d \geq 1$ input variables, $\mathbf{x} = \{x_1, \dots, x_d\}$, with support $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^d$. The prototype of this approach, Sobol's method is a generalization of the analysis of variance (ANOVA) to $d > 2$ independent input variables and decomposes y , as sum of elementary functions of zeroth-, first-, second-, up to d th-order. This independence assumption is mathematically convenient but may not be borne out of the causal or correlational relationships between the x 's. This paper is concerned with variance-based sensitivity analysis (SA) for correlated input variables, for example, multivariate dependencies in a posterior parameter distribution. We use high-dimensional model representation (HDMR) of Li et al. (2010, <https://doi.org/10.1021/jp9096919>), Li and Rabitz (2012, <https://doi.org/10.1007/s10910-011-9898-0>) and replace Sobol's elementary functions with so-called component functions with unknown expansion coefficients to disentangle the structural, correlative and total contribution of input factors. We contrast the default HDMR methodology with cubic B-splines and sequential coefficient estimation against its successor, HDMR_{ext} of Li and Rabitz (2012, <https://doi.org/10.1007/s10910-011-9898-0>), which uses polynomial component functions with an extended orthonormalized basis. Benchmark experiments confirm that HDMR and HDMR_{ext} parse out the structural and correlative contributions of input factors to the model output and infer an optimal experimental design with parameter correlation. Our last study applies HDMR_{ext} to probabilistic SA of a watershed model. The multivariate posterior parameter distribution supports model emulation and yields sensitivity indices that pertain to measured discharge data.

1. Introduction

Sensitivity analysis (SA) is commonly used to ascertain the relative importance of input variables, $\mathbf{x} = \{x_1, \dots, x_d\}$, in determining the simulated output, $\mathbf{y} = \{y_1, \dots, y_n\}$, of some vector-valued function or model, $\mathbf{f}(\mathbf{x})$ (Saltelli et al., 2010). A closely related practice is uncertainty analysis, which is concerned with quantifying the confidence (prediction) limits of the simulated output, y_i ; $t = (1, \dots, n)$. This usually involves the use of training data record, $\tilde{\mathbf{y}} = \{\tilde{y}_1, \dots, \tilde{y}_n\}$, and requires prior assumptions about the nature and distribution of the residuals, $\epsilon_t = \tilde{y}_t - y_t$, and the sources of modeling errors (Beven & Binley, 1992; Liu & Gupta, 2007; Vrugt et al., 2005). Thus, uncertainty analysis quantifies the probability that a certain event (or range of events) will take place, whereas SA does not tell us anything about how likely a certain outcome of the model is. This essay is concerned with a general description of sensitivity across the confidence intervals of the simulated output. This so-called probabilistic SA (Oakley & O'Hagan, 2004) quantifies the sensitivity of simulated and/or forecasted event probabilities for a given multivariate probability distribution of the input variables. Our methodology relies on the analysis of covariance (ANCOVA) to preserve the multivariate character of the $(\mathbf{x}, \mathbf{f}(\mathbf{x}))$ -relationship.

SA is an essential step in model development, model calibration and quality assurance (Saltelli et al., 2020). In the past decades, many different SA methods have been developed and used in the mathematical and applied literature. Of these approaches, variance-based methods are particularly attractive because of their innate ability to characterize sensitivity over the entire prior ranges of the input variables and capacity to differentiate between the marginal, joint and total effects of x_1, \dots, x_d (Cukier et al., 1973; McKay, 1995; Razavi & Gupta, 2016; Sobol', 1990). The prototype of this approach, Sobol' (1990) method, was originally presented in Russian. In the English reprint, Sobol' (1993) showed that the output, $y = f(\mathbf{x})$, of a scalar-valued square-integrable function, $f \in L^2(\mathbb{K}^d)$, on the d -dimensional unit cube, $\mathbb{K}^d = \{\mathbf{x} | 0 \leq x_i \leq 1; i = 1, \dots, d\}$, can be uniquely decomposed into summands of elementary functions, $f_1(x_1), f_{ij}(x_i, x_j), \dots, f_{12\dots d}(x_1, x_2, \dots, x_d)$, under orthogonality constraints, to yield

Writing – original draft: Yifu Gao
Writing – review & editing: Yifu Gao,
Jasper A. Vrugt

$$y = f(\mathbf{x}) = f_0 + \sum_{i=1}^d f_i(x_i) + \sum_{i=1}^{d-1} \sum_{j=i+1}^d f_{ij}(x_i, x_j) + \sum_{i=1}^{d-2} \sum_{j=i+1}^{d-1} \sum_{k=j+1}^d f_{ijk}(x_i, x_j, x_k) + \dots + f_{12\dots d}(x_1, x_2, \dots, x_d) + \epsilon, \quad (1)$$

where f_0 signifies the expected (mean) value, $f_0 = \mathbb{E}[f(\mathbf{x})]$, of the model output and ϵ is the residual. This residual is assumed to follow a normal distribution, $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$, with constant (unexplained) variance, σ_ϵ^2 . The first-order functions, $f_i(x_i)$, characterize the individual effects of the input variables on the simulated model output. The second-, $f_{ij}(x_i, x_j)$, third-, $f_{ijk}(x_i, x_j, x_k)$, up to d th-order effects, $f_{12\dots d}(x_1, x_2, \dots, x_d)$, characterize the contribution of groups of two, three, up to all input variables combined to y . These functions account for interactions between the input variables that contribute partially to the model output. Equation 1 means that

$$y = f(\mathbf{x}) = f_0 + \sum_{i=1}^{n_1} f_i(x_i) + \sum_{1 \leq i < j \leq d}^{n_2} f_{ij}(x_i, x_j) + \sum_{1 \leq i < j < k \leq d}^{n_3} f_{ijk}(x_i, x_j, x_k) + \dots + f_{12\dots d}(x_1, x_2, \dots, x_d) + \epsilon, \quad (2)$$

with $n_1 = d$, main effects, $n_2 = d(d-1)/2$, second-order effects, $n_3 = d(d-1)(d-2)/6$, third-order effects, and so forth. The total number of additive terms (summands) in the expression above equals 2^d .

Unknown to Sobol' (1993) at the time, the decomposition of Equation 2 was derived decades earlier by Hoeffding (1948) as a generalization of the analysis of variance (ANOVA) to $d > 2$ independent variables. Functional decomposition of the variance of the model output, y , in Equation 2 produces the following identity (Hoeffding, 1948; Sobol', 1990)

$$\text{Var}[y] = \sum_{i=1}^{n_1} \text{Var}[f_i(x_i)] + \sum_{1 \leq i < j \leq d}^{n_2} \text{Var}[f_{ij}(x_i, x_j)] + \sum_{1 \leq i < j < k \leq d}^{n_3} \text{Var}[f_{ijk}(x_i, x_j, x_k)] + \dots + \text{Var}[f_{12\dots d}(x_1, x_2, \dots, x_d)] + \sigma_\epsilon^2, \quad (3)$$

If we divide all terms by $\text{Var}[y]$, we yield Sobol' sensitivity indices

$$\sum_{i=1}^{n_1} S_i(x_i) + \sum_{1 \leq i < j \leq d}^{n_2} S_{ij}(x_i, x_j) + \sum_{1 \leq i < j < k \leq d}^{n_3} S_{ijk}(x_i, x_j, x_k) + \dots + S_{12\dots d}(x_1, \dots, x_d) = 1, \quad (4)$$

pending the assumption that $\sigma_\epsilon^2 \ll \text{Var}[y]$. The S_i 's are the individual (main) effects of the input variables on the simulated output and the sensitivity indices with more than one subscript are the interactions. As the order of the indices in the subscript of the interaction terms is inconsequential, we follow convention and sort the indices from low to high.

A useful related measure is the total-order or total-effect sensitivity index (Homma & Saltelli, 1996)

$$S_i^T(x_i) = S_i(x_i) + \sum_{i \neq j} S_{ij}(x_i, x_j) + \sum_{i \neq j \neq k} S_{ijk}(x_i, x_j, x_k) + \dots + S_{12\dots d}(x_1, \dots, x_d), \quad (5)$$

and includes all Sobol' indices of Equation 4 with index i in the subscript. The total-order sensitivity index equals the fraction of variance that would remain on average if x_i is left to vary over its prior range while all other input variables are fixed (Puy et al., 2021). Note that $\sum_{i=1}^d S_i^T(x_i) \geq 1$ as interaction terms, say, $S_{ij}(x_i, x_j)$, count in both S_i^T and S_j^T . Most mathematical functions do not admit a closed-form solution for the Sobol' indices, and, therefore, we must resort to a numerical procedure to estimate the sensitivity indices of Equation 4. This includes the (quasi)-Monte Carlo approaches of Sobol' (2001) and Owen (2005), Fourier amplitude sensitivity testing (Cukier et al., 1973), random balance designs (Tarantola et al., 2006) and variogram analysis of response surfaces (Razavi & Gupta, 2016) and involve repeated evaluation of the model output $f(\mathbf{x})$ for $\mathbf{x} \in \mathcal{X}$. Note that the last three methods solve only for the main and/or total-order indices of the input factors.

Sobol's decomposition is unique only under orthogonality constraints. If u is a subset of selected dimensions of the input variables, $u \subseteq \{1, \dots, d\}$, then Equation 2 is synonymous to a d -way ANOVA-expansion of $f(\mathbf{x})$ under one of two constraints (Chastaing et al., 2015)

$$\int_0^1 f_u(\mathbf{x}_u) dx_r = 0 \quad \text{for all } r \in u \quad (6a)$$

$$\int_0^1 f_u(\mathbf{x}_u) f_v(\mathbf{x}_v) dx = 0 \quad \text{for all } u, v \subseteq \{1, \dots, d\}; u \neq v, \quad (6b)$$

Thus, the expected value of $f_u(\mathbf{x}_u)$ with respect to each input variable of \mathbf{x}_u should be zero or, alternatively, $\mathbb{E}[f_u(\mathbf{x}_u)f_v(\mathbf{x}_v)] = 0$, thus the inner product, $\langle f_u(\mathbf{x}_u), f_v(\mathbf{x}_v) \rangle$ of pairwise elementary functions should be zero. Linear independence of two random variables, X and Y , implies that the variance of their sum, $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$. Thus, the orthogonality constraint of Equation 6 eliminates the covariances between the summands of elementary functions. Then Equations 3 and 4 are valid and unique.

The independence requirement of x_1, \dots, x_d is mathematically convenient but may not be borne out of the causal or correlational relationships between the input variables. Such association is not uncommon in hydrologic modeling and may be imposed by constraints on the input variables or by the fact that experimental data, $\tilde{\mathbf{y}} = \{\tilde{y}_1, \dots, \tilde{y}_n\}$, and/or expert judgment are used (Mara et al., 2015). Indeed, correlation among x_1, \dots, x_d is almost guaranteed if the model output originates from the posterior probability density function, $p(\mathbf{x}|\tilde{\mathbf{y}})$, of the input variables (Gomes et al., 2017; Pan et al., 2009; Schoups et al., 2010; Song et al., 2015; Vrugt et al., 2008; H. Wang, Gong, et al., 2020). This induced structure in the $(\mathbf{x}, f(\mathbf{x}))$ -relationship violates the expansion of Hoeffding (1948) or Sobol' (1993) in Equation 2 and demands an ANCOVA-based model decomposition. Several publications have discussed the importance of parameter correlation in hydrologic SA (Song et al., 2015; Yuan et al., 2015), but most applications of global SA in hydrologic modeling continue to rely on parameter independence (A. Wang & Solomatine, 2019; A. Wang, Pianosi, et al., 2020; Zelelew & Alfreksen, 2013). A few studies have considered parameter correlation in hydrologic SA but the dependent parameters are (a) either lumped together (Khare et al., 2013) or (b) assumed to follow a convenient multivariate normal distribution (Ahn, 1996; Manache & Melching, 2008; Pan et al., 2011; Zhu, 2012) or isoprobabilistic Nataf transformation (Do & Razavi, 2020). The so-obtained sensitivity indices do not unify with Sobol' indices in Equation 4 or in the case of the gSTAR-VARS method of Do and Razavi (2020) provide only total-order effects.

In this paper, we are concerned with variance-based SA of hydrologic models in the presence of correlated input variables, for example, multivariate dependencies among the parameters of a posterior distribution. We use high dimensional model representation or HDMR (Li & Rabitz, 2012; Li et al., 2010) and replace the elementary functions in Sobol's decomposition of Equation 2 by so-called component functions with linear expansion coefficients as unknown parameters. If the component functions satisfy hierarchical orthogonality, this meta-modeling approach disentangles exactly the structural and correlative contributions of individual and groups of input variables to the model output. The resulting sensitivity indices are a generalization of Sobol' indices in Equation 4. The HDMR methodology has been used in earlier publications in the hydrologic literature to prioritize and/or rank input factors (Bennett & Fentie, 2017), quantify interaction terms of the input factors (Borgonovo et al., 2017) and assess the direction of change in model outputs (A. Wang, Pianosi, et al., 2020). But these studies are limited to independent input variables, do not necessarily satisfy hierarchical orthogonality of the component functions and/or use a first-order model decomposition only. We use first-, second- and third-order HDMR model decomposition and examine in detail how the structural, correlative and total sensitivity indices depend on soil and watershed wetness. Furthermore, we investigate the impact of parameter correlation on experimental design and demonstrate the use of the HDMR decomposition for model emulation.

The HDMR method arguably is one of the most advanced SA methods, but by no means the only variance-based approach that can handle correlation between the input factors. These approaches simplify ANCOVA-based function decomposition to a correlated and uncorrelated contribution (McKay, 1997; Xu, 2013; Xu & Gertner, 2008a, 2008b) or main effects and total sensitivity indices (Kucherenko et al., 2012, 2017), made possible in part by application of Sobol's method to orthogonalized input variables derived from grouping (Jacques et al., 2006; Saltelli & Tarantola, 2002), Gram-Schmidt projection (Chastaing et al., 2015; Mara & Tarantola, 2012) and isoprobabilistic transformation (Mara et al., 2015; Tarantola & Mara, 2017). HDMR decomposes the variance-covariance structure of the input-output relationship into a structural and correlative contribution at first-, second-, and/or higher orders. This general-purpose method admits samples of the posterior distribution of the input variables for probabilistic SA of the $(\mathbf{x}, f(\mathbf{x}))$ -relationship.

The remainder of this paper is organized as follows. Section 2 summarizes the problem at hand. Section 3 presents rationales and theory of HDMR and discusses its default implementation and a variant with extended bases, HDMR_{ext}. This is followed in Section 4 by application of both methods to four case studies of increasing complexity. The first two studies illustrate and benchmark the performance of HDMR and HDMR_{ext} for dependent and independent input variables by application to a simple linear model and the well-known Ishigami function (Ishigami & Homma, 1990). The next two case studies demonstrate the power and usefulness of HDMR_{ext} in

hydrologic modeling using SA of the soil water characteristic of van Genuchten (1980) and the rainfall-discharge relationship using hmodel (Schoups et al., 2010). In these last two studies, we analyze traces of the structural and correlative sensitivity indices and examine in detail the performance of the HDMR_{ext} emulator as a cheap surrogate of the forward model. Finally, Section 5 concludes this paper with a summary and discussion of our main findings. Specifically, we address the strengths and weaknesses of the HDMR and HDMR_{ext} , and provide some suggestions for future developments.

2. Problem Statement

Let's consider a real-world system, \mathfrak{S} , whose future state, behavior and regularities may be described with explanatory laws, for example, those most fundamental and celebrated in physics. Our model, $\mathbf{Y} \leftarrow \mathbf{f}(\mathbf{x}, \mathbf{\Omega})$, scales up the physical laws to the domain of \mathfrak{S} , simulates its spatiotemporal evolution and returns a matrix, \mathbf{Y} , of simulated values of different quantities of interest. The d -vector, $\mathbf{x} = \{x_1, \dots, x_d\}$, specifies the system and/or material properties of \mathfrak{S} and the array, $\mathbf{\Omega}$, consists of all other variables (constant or not) believed to govern, by causality, the variables of \mathbf{Y} using the evolution rules of nature expressed in mathematical form. Generally, \mathbf{f} is not available in closed form but we assume that it can be evaluated at any point, $\mathbf{x} = \{x_1, \dots, x_d\} \in \mathcal{X} \subseteq \mathbb{R}^d$.

As most real-world systems have an intractable complexity, the dimensionality of the input space, $(\mathbf{x}, \mathbf{\Omega})$, may grow very large to accurately portray the spatiotemporal evolution of \mathfrak{S} . This large model complexity will cloud the relationship between model inputs and outputs. SA will help determine which input variables exert the largest control on the simulated output, y , and which variables (parameters) have a negligible impact, thus, is an essential ingredient of model building and quality assurance (Saltelli et al., 2020). This serves many different purposes among which model simplification, model development, model calibration and uncertainty analysis and/or reduction are the most important applications.

This paper addresses a critical limitation of commonly used SA methods in hydrology and is concerned with global SA in the presence of correlated input variables. Such linear and/or nonlinear dependencies violate the variance summation used in Sobol's method and necessitate application of a mathematically much more rigorous but arguably more difficult ANCOVA-based model decomposition. Before doing so, we first clarify our mathematical notation. Without loss of generality, we focus in this paper on the vector of model parameters, \mathbf{x} , and consider only a single output, y . Thus, we suppress the array $\mathbf{\Omega}$ in our subsequent notation and write, $y = f(\mathbf{x}) : \mathbb{R}^d \rightarrow \mathbb{R}^1$, for the scalar-valued form of the model with respect to \mathbf{x} . The entries of the d -vector, $\mathbf{x} = \{x_1, \dots, x_d\}$, are referred to as input variables, input factors or parameters in the remainder of this paper. After computation of the model output, $y = f(\mathbf{x})$, the input variables are normalized to the unit d -dimensional cube, $\mathbb{K}^d = \{\mathbf{x} \mid 0 \leq x_i \leq 1 \text{ for all } i = 1, \dots, d\}$, prior to SA.

In this paper, we are concerned with a rigorous variance-based description of the structural and correlative contribution of hydrologic input variables to the model output and not with computational efficiency. This is inconsequential for our current case studies, but can pose a major limitation for high-dimensional and/or CPU-intensive hydrologic models. We will address the computational requirements of the HDMR methodology in passing. Although we use Sobol's method as main reference, we reiterate that this paper is not about a comparison of variance-based methods but rather about (probabilistic) SA in the presence of dependent input variables.

3. High Dimensional Model Representation

In this section, we present a brief review of the main theory and ingredients of the HDMR methodology of Li et al. (2010) and describe its most up to date implementation with extended bases, so-called HDMR_{ext} , from Li and Rabitz (2012).

3.1. Main Theory

The HDMR method is an extension of Sobol's method to correlated input variables. HDMR is rooted in statistical theory and uses variance/covariance-based decomposition of the model output to determine the structural and correlative contributions of each input variable separately and combinations thereof. Projector operator theory simplifies the learning of the input-output relationships to polynomial complexity.

HDMR builds on the model output decomposition in Equation 2 and replaces the elementary functions, $f_i(x_i), f_{ij}(x_i, x_j), \dots, f_{12\dots d}(x_1, x_2, \dots, x_d)$, with so-called component functions with linear expansion coefficients as unknown

parameters. Suppose that we have a total of $n_{123} = n_1 + n_2 + n_3$ first-, second-, and third-order component functions. We can simplify Equation 2 to read

$$y = f(\mathbf{x}) = f_0 + \sum_{u=1}^{n_{123}} f_u + \epsilon, \quad (7)$$

where the subscript u signifies the index of the component function, rather than its order as in Equation 2. Thus, f_1, \dots, f_d , correspond to the first-order component functions, $f_{d+1}, f_{d+2}, \dots, f_{d+d(d-1)/2}$, signify the second-order component functions and, $f_{d+d(d-1)/2+1}, \dots, f_{d+d(d-1)/2+d(d-2)/6}$, equal the third-order component functions. According to the functional classification of Kucherenko et al. (2011), component functions up to the third-order will usually suffice for describing the $(\mathbf{x}, f(\mathbf{x}))$ -relationship of systems with not equally important variables (type A) and systems with equally important variables and dominant low order terms (type B). Most physical systems are of these two types (e.g. Falchi et al., 2018; Li & Rabitz, 2012; Li et al., 2010; Rabitz & Aliş, 1999; Ratto et al., 2007; Shereena & Rao, 2019; H. Wang et al., 2017; Ziehn & Tomlin, 2008). The last class of functions with equally important variables and dominant interaction terms, type C identified by Kucherenko et al. (2011), may demand fourth and/or higher-order component functions. This may introduce problems with the well-posedness of the HDMR functional decomposition, unless a very large number of input-output samples is used. Note that the sum of the component functions at the right hand side of Equation 7 may serve as a cheap surrogate model (emulator) of the original system model, $f(\mathbf{x})$. The emulator provides an approximation of $f(\mathbf{x}) - f_0$ up to the residual, $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$. In practice, we approximate σ_ϵ^2 by the sample variance, s_ϵ^2 , of the residuals of the surrogate model. The change in subscript notation of the component functions suits variance decomposition of the simulated output

$$\text{Var}[y] = \mathbb{E}[(y - \mathbb{E}(y))^2] = \int_{\mathbb{K}^d} w(\mathbf{x})(y - f_0)^2 \, d\mathbf{x}, \quad (8)$$

where $w(\mathbf{x})$ denotes the probability density function of the d -variate prior distribution of the input variables, \mathbf{x} . If we substitute Equation 7 into Equation 8 we yield the following expression (Li et al., 2010)

$$\text{Var}[y] = \int_{\mathbb{K}^d} w(\mathbf{x}) \left(\sum_{u=1}^{n_{123}} f_u + \epsilon \right)^2 \, d\mathbf{x}, \quad (9)$$

which is equivalent to

$$\text{Var}[y] = \int_{\mathbb{K}^d} w(\mathbf{x}) \left(\sum_{u=1}^{n_{123}} f_u \right)^2 \, d\mathbf{x} + 2 \int_{\mathbb{K}^d} w(\mathbf{x}) \sum_{u=1}^{n_{123}} (f_u \cdot \epsilon) \, d\mathbf{x} + \mathbb{E}[\epsilon^2], \quad (10)$$

where \mathbb{K}^d equals the unit cube of the normalized input variables. Least squares determination of the different f_u 's enforces orthogonality between the component functions and residuals, $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$. As a result, the inner product of each f_u and ϵ equals zero. Thus, the second term dissipates

$$\int_{\mathbb{K}^d} w(\mathbf{x})(f_u \cdot \epsilon) \, d\mathbf{x} = 0. \quad (11)$$

and Equation 10 simplifies to Li et al. (2010)

$$\text{Var}[y] = \int_{\mathbb{K}^d} w(\mathbf{x}) \left(\sum_{u=1}^{n_{123}} f_u \right)^2 \, d\mathbf{x} + \mathbb{E}[\epsilon^2]. \quad (12)$$

This results in the following formulation for the variance of the model output, y , in Equation 7

$$\text{Var}[y] = \sum_{u=1}^{n_{123}} \text{Var}[f_u] + \sum_{u=1}^{n_{123}} \text{Cov} \left[f_u, \sum_{m \neq u} f_m \right] + \mathbb{E}[\epsilon^2]. \quad (13)$$

The variance of the model output equals the sum of the variances and covariances of the component functions. If we normalize the variance and covariance contribution of each component function with the variance of the model output

$$S_u^a = \frac{\text{Var}[f_u]}{\text{Var}[y]} \quad \text{and} \quad S_u^b = \frac{\text{Cov}\left[f_u, \sum_{m \neq u}^{n_{123}} f_m\right]}{\text{Var}[y]}, \quad (14)$$

we yield the *structural*, S_u^a , and *correlative*, S_u^b , sensitivity of each component function, f_u , and, thus, input variable or combinations thereof. The total sensitivity index, S_u , of the u th component function, f_u , now simply equals the sum of its structural and correlative sensitivity indices to yield (Li et al., 2010)

$$S_u = S_u^a + S_u^b. \quad (15)$$

For a square-integrable function and proper component functions, Equation 7 should approximate well the simulated output and the variance, σ_ϵ^2 , of the residuals, should be small. Then, $\mathbb{E}[\epsilon^2] \approx 0$ and, thus

$$\sum_{u=1}^{n_{123}} S_u \approx \frac{\text{Var}[y]}{\text{Var}[y]} = 1, \quad (16)$$

the sum of the total sensitivity indices of the component functions should amount to one. When the input variables are independent, the correlative contribution of the component functions will be zero by definition and Equation 13 simplifies to Sobol' variance decomposition

$$\text{Var}[y] = \sum_{u=1}^{n_{123}} \text{Var}[f_u] + \mathbb{E}[\epsilon^2], \quad (17)$$

and the total sensitivity index, S_u , of each component function will then amount to its structural sensitivity index, S_u^a , only. Thus, for independent input variables, HDMR reduces to Sobol's method.

We are left with the choice of the component functions of Equation 7. Unfortunately, the model decomposition in Equation 7 does not admit a convenient closed-form solution for the mathematical form of the component functions (Khorashadi Zadeh et al., 2017; Li et al., 2001). We must therefore prescribe a suitable functional form of the component functions. Li et al. (2010) used uniform cubic B-splines for the first-, second- and third-order component functions, $f_i(x_i)$, $f_{ij}(x_i, x_j)$, and $f_{ijk}(x_i, x_j, x_k)$, respectively, of Equation 2

$$f_i(x_i) = \sum_{r=0}^{m+2} \alpha_r^i B_{r,3}(x_i) \quad (18a)$$

$$f_{ij}(x_i, x_j) = \sum_{r=0}^{m+2} \sum_{s=0}^{m+2} \beta_{rs}^{ij} B_{r,3}(x_i) B_{s,3}(x_j) \quad (18b)$$

$$f_{ijk}(x_i, x_j, x_k) = \sum_{r=0}^{m+2} \sum_{s=0}^{m+2} \sum_{t=0}^{m+2} \gamma_{rst}^{ijk} B_{r,3}(x_i) B_{s,3}(x_j) B_{t,3}(x_k), \quad (18c)$$

where m equals the number of so-called Bézier curves of each cubic B-spline, α , β , and γ are unknown expansion coefficients of the component functions and $B_{r,3}(x_i)$, $B_{s,3}(x_j)$, and $B_{t,3}(x_k)$ are defined on the unit intervals of \mathbb{K}^d by recursion using the Cox-de Boor algorithm (Cox, 1972; De Boor, 1972). Thus, for $m = 4$, the number of control points of each cubic B-spline is equal to $m + 3 = 7$.

If we substitute Equations 18a–18c into Equation 2 we yield the following expression for the model output

$$y = f(\mathbf{x}) = f_0 + \sum_{i=1}^{n_1} \left\{ \sum_{r=0}^{m+2} \alpha_r^i B_{r,3}(x_i) \right\} + \sum_{1 \leq i < j \leq d}^{n_2} \left\{ \sum_{r=0}^{m+2} \sum_{s=0}^{m+2} \beta_{rs}^{ij} B_{r,3}(x_i) B_{s,3}(x_j) \right\} + \sum_{1 \leq i < j < k \leq d}^{n_3} \left\{ \sum_{r=0}^{m+2} \sum_{s=0}^{m+2} \sum_{t=0}^{m+2} \gamma_{rst}^{ijk} B_{r,3}(x_i) B_{s,3}(x_j) B_{t,3}(x_k) \right\} + \epsilon, \quad (19)$$

which has a total of $d(m + 3)$ unknown α -coefficients (first order), $d(d - 1)(m + 3)^2/2$ unknown β -coefficients (second order) and $d(d - 1)(d - 2)(m + 3)^3/6$ unknown γ -coefficients (third order). The total number of expansion coefficients, l , of the HDMR functional decomposition in Equation 19 is now equal to

$$l = d(m + 3) + \frac{1}{2}d(d - 1)(m + 3)^2 + \frac{1}{6}d(d - 1)(d - 2)(m + 3)^3. \quad (20)$$

and articulates a linear, quadratic and cubic growth of the number of expansion coefficients of the first-, second-, and third-order component functions, respectively, with the number of input variables d and control points, $m + 3$, of the B-splines. To guarantee a well-posed inverse problem, the number of input-output samples N should exceed the number of expansion coefficients, l . In general, the larger the quotient of N by l , the more robust the estimates of α , β , and γ will be and the smaller the confidence intervals of the HDMR sensitivity indices. In our case studies, the ratio N/l is at least equal to seven but often considerably larger.

As the variance decomposition of Equation 7 is additive and the basis functions of Equations 18a–18c are valid we can resort to regularized least squares to simultaneously determine the expansion coefficients, α , β , and γ , of all n_{123} component functions from a collection of N different (\mathbf{x}, y) -data pairs. As the number of expansion coefficients grows rapidly with dimensionality, d , of the space, \mathbb{K}^d , of input variables, Li et al. (2010) recommends using backfitting instead. This method starts at the first-order and determines sequentially (one after the next) the expansion coefficients of each component function. A F -test is used to discard insignificant component functions of Equation 7 and reduce the required number of expansion coefficients, l , as much as possible. If RSS_0 signifies the residual sum of squares of the emulator, $y = f_0 + \sum_{i=1}^{d-1} f_i(x_i)$, with $l_1 = (d - 1)(m + 3)$ unknown expansion coefficients according to Equation 18a and RSS_1 is the same quantity for the surrogate model, $y = f_0 + \sum_{i=1}^d f_i(x_i)$, expanded with $f_d(x_d)$ and $l_0 = d(m + 3)$ expansion coefficients, then the F -statistic

$$F = \frac{(RSS_0 - RSS_1)/(l_1 - l_0)}{RSS_1/(N - l_1)}, \quad (21)$$

will determine whether the first-order component function of x_d is significant or not. We must reject the null hypothesis, “ $\mathcal{H}_0 : f_d(x_d)$ is insignificant,” if the F -statistic exceeds the critical value, $F_{\mathcal{F}}^{-1}(p_\alpha | l_1 - l_0, N - l_1)$, of the F -distribution at confidence level, $p_\alpha \in (0, 1)$, and $l_1 - l_0$, and $N - l_1$ degrees of freedom. One can use forward selection (starting from f_0) or backward elimination (starting from $f_0 + \sum_{i=1}^d f_i(x_i)$) to determine the significant first-order component functions. Then, a similar procedure can be used for the second-order component functions, with significant first-order component functions fixed and so forth. Figure S1 in Supporting Information S1 summarizes the different steps of the HDMR methodology of Li et al. (2010). Interested readers are referred to this publication for more information.

The model decomposition of Equation 2 considers all possible orders of the component functions up to $f_{12\dots d}(x_1, \dots, x_d)$. As the higher order component function usually contribute little to the simulated output, y , the HDMR functional decomposition in Equation 19 uses only first-, second- and third-order component functions. If deemed appropriate, we can discard the third-order component functions and work with a second- or possibly even first-order model decomposition instead. We can evaluate the overall performance of the HDMR decomposition with different order component functions to determine a suitable maximum order $\delta \in \mathbb{N}_+$ of the component functions. Unless stated differently, we use $m = 2$ Bézier curves for the cubic B-splines and implement forward selection of the HDMR component functions using $p_\alpha = 0.99$.

3.2. Extended Bases

Sequential determination of the component functions of each given order with backfitting in the HDMR method may be computationally appealing but has important side effects that become more and more debilitating with increasing dimensionality, d , and degree of correlation of the input variables. Under those circumstances, backfitting may suffer convergence problems in pursuit of the optimum values of the expansion coefficients and does not guarantee a minimum variance of the residuals. Then the total sensitivity indices, S_{μ^*} , of the n_{123} component functions may not sum to one. A more profound problem with the HDMR implementation is that the cubic B-splines may not satisfy hierarchical orthogonality. This provokes further problems with the variance decomposition of the model output.

To guarantee an exact decomposition of the variance of the model output, y , Li and Rabitz (2012) introduced an alternate implementation of the HDMR methodology with extended bases. This method, abbreviated herein to $HDMR_{\text{ext}}$, enforces hierarchical orthogonality between the component functions. Furthermore, the coefficients, α , β , and γ of the first-, second-, and third-order component functions are simultaneously determined using least squares regression. These modifications guarantee an exact description of the structural and correlative sensitivity indices of the input variables, x_1, \dots, x_d .

Hierarchical orthogonality of the component functions is guaranteed if we satisfy a so-called relaxed vanishing condition (Hooker, 2007)

$$\int_0^1 w_u(\mathbf{x}_u) f_u(\mathbf{x}_u) dx_i = 0 \quad \text{for all } u \subseteq \{1, \dots, d\} \text{ and } i \in u, \quad (22)$$

where u is a subset of elements of the superset $U = \{1, \dots, d\}$, \mathbf{x}_u denote the dimensions u of the input vector and $w_u(\mathbf{x}_u)$ signifies the probability density function (pdf) of \mathbf{x}_u . For a second-order component function, the vanishing condition of Equation 22 dictates that $f_{ij}(x_i, x_j)$ should be orthogonal to its lower order component functions, $f_i(x_i)$ and $f_j(x_j)$. With independent input variables, the multivariate pdf $w_u(\mathbf{x}_u)$ equals a $|u|$ -variate uniform distribution with constant density for all $\mathbf{x}_u \in \mathbb{K}^{|u|}$, and Equation 22 reduces to Equation 2 of Sobol' (2001).

To satisfy the vanishing condition of Equation 22, Li and Rabitz (2012) resort to the family of orthogonal polynomial functions

$$\phi_1(x_i) = a_1 x_i + a_0 \quad \text{degree } p = 1 \quad (23a)$$

$$\phi_2(x_i) = b_2 x_i^2 + b_1 x_i + b_0 \quad \text{degree } p = 2 \quad (23b)$$

$$\phi_3(x_i) = c_3 x_i^3 + c_2 x_i^2 + c_1 x_i + c_0 \quad \text{degree } p = 3, \quad (23c)$$

where the coefficients, a , b , and c , are determined by applying Gram-Schmidt orthonormalization to the respective basis, x^3 , x^2 , x and 1 of the polynomials. This projection operator constructs an orthonormal basis for the polynomial functions on the unit interval of x with respect to an arbitrary weighting function. The component functions of Equation 7 are now equal to the sum of linear multiples of the orthonormalized polynomial functions of degrees 1 to p to yield

$$f_i(x_i) = \sum_{r=1}^p \alpha_r^{(ii)} \phi_r(x_i) \quad (24a)$$

$$f_{ij}(x_i, x_j) = \sum_{r=1}^p [\alpha_r^{(ij)i} \phi_r(x_i) + \alpha_r^{(ij)j} \phi_r(x_j)] + \sum_{r=1}^p \sum_{s=1}^p \beta_{rs}^{(ij)ij} \phi_r(x_i) \phi_s(x_j) \quad (24b)$$

$$f_{ijk}(x_i, x_j, x_k) = \sum_{r=1}^p [\alpha_r^{(ijk)i} \phi_r(x_i) + \alpha_r^{(ijk)j} \phi_r(x_j) + \alpha_r^{(ijk)k} \phi_r(x_k)] + \sum_{r=1}^p \sum_{s=1}^p [\beta_{rs}^{(ijk)ij} \phi_r(x_i) \phi_s(x_j) + \beta_{rs}^{(ijk)ik} \phi_r(x_i) \phi_s(x_k)] + \beta_{rs}^{(ijk)jk} \phi_r(x_j) \phi_s(x_k) + \sum_{r=1}^p \sum_{s=1}^p \sum_{t=1}^p \gamma_{rst}^{(ijk)ijk} \phi_r(x_i) \phi_s(x_j) \phi_t(x_k), \quad (24c)$$

where the expansion coefficients, α , β , and γ , may be estimated from a large collection of (\mathbf{x}, y) -data pairs using least squares regression. Unlike the cubic B-splines in HDMR, the component functions of HDMR_{ext} are written as sum of linear multiples of orthonormalized polynomial functions of degrees 1 to p . This is what Li and Rabitz (2012) refers to as extended bases and is required to satisfy the vanishing condition of Equation 22. The use of extended bases has implications for our index notation of the coefficients. The symbol(s) in the parenthesis of each superscript of α , β and γ convey the component function. The symbols that are listed next in superscript correspond to the indices of the input vector.

Suppose that all component functions have the same polynomial degree, p . This is not a requirement of our MATLAB implementation of HDMR_{ext} but simplifies the mathematical description. Then, if we substitute Equations 24a–24c into Equation 2 we yield the following expression for the model output

$$\begin{aligned}
 y = f(\mathbf{x}) = & f_0 + \sum_{i=1}^d \sum_{r=1}^p \alpha_r^{(i)i} \phi_r(x_i) \\
 & + \sum_{1 \leq i < j \leq d} \left\{ \sum_{r=1}^p [\alpha_r^{(ij)i} \phi_r(x_i) + \alpha_r^{(ij)j} \phi_r(x_j)] + \sum_{r=1}^p \sum_{s=1}^p \beta_{rs}^{(ij)ij} \phi_r(x_i) \phi_s(x_j) \right\} \\
 & + \sum_{1 \leq i < j < k \leq d} \left\{ \sum_{r=1}^p [\alpha_r^{(ijk)i} \phi_r(x_i) + \alpha_r^{(ijk)j} \phi_r(x_j) + \alpha_r^{(ijk)k} \phi_r(x_k)] \right. \\
 & + \sum_{r=1}^p \sum_{s=1}^p [\beta_{rs}^{(ijk)ij} \phi_r(x_i) \phi_s(x_j) + \beta_{rs}^{(ijk)ik} \phi_r(x_i) \phi_s(x_k) \\
 & \left. + \beta_{rs}^{(ijk)jk} \phi_r(x_j) \phi_s(x_k)] + \sum_{r=1}^p \sum_{s=1}^p \sum_{t=1}^p \gamma_{rst}^{(ijk)ijk} \phi_r(x_i) \phi_s(x_j) \phi_t(x_k) \right\} + \epsilon.
 \end{aligned} \tag{25}$$

According to Equation 24 we have $d \cdot p$, $d(d-1)p$, and $d(d-1)(d-2)p/2$ different α -coefficients in the first-, second-, and third-order component functions, respectively. Similarly, we have $d(d-1)p^2/2 + d(d-1)(d-2)p^2/2$ different β 's in the second- and third-order component functions. Furthermore, the third-order component functions also have $d(d-1)(d-2)p^3/6$ γ -coefficients. This brings the total number of expansion coefficients of Equation 25 to

$$l = \underbrace{d \cdot p}_{\delta=1} + \underbrace{\frac{1}{2}d(d-1)(2p+p^2)}_{\delta=2} + \underbrace{\frac{1}{6}d(d-1)(d-2)(3p+3p^2+p^3)}_{\delta=3}, \tag{26}$$

where δ signifies the maximum order of the HDMR_{ext} decomposition. Thus, the complexity of the HDMR_{ext} emulator is governed by the dimensionality, d , of the input space \mathbb{K}^d and the order δ and polynomial degree p of the component functions.

We can write Equation 25 as a vector inner product

$$\Phi(\mathbf{x})^\top \mathbf{c} = f(\mathbf{x}) - f_0 \tag{27}$$

of a $1 \times l$ design vector, $\Phi(\mathbf{x})^\top$, with orthonormalized polynomial functions of Equation 23 (and products thereof) evaluated at their respective entries of \mathbf{x} and arranged in order of their appearance in Equation 25

$$\begin{aligned}
 \Phi(\mathbf{x}) = & [\phi_1(x_1) \dots \phi_1(x_2) \dots \phi_p(x_d) \mid \phi_1(x_1) \phi_1(x_2) \dots \phi_p(x_1) \phi_p(x_2) \dots \phi_p(x_1) \phi_p(x_3) \\
 & \dots \phi_p(x_{d-1}) \phi_p(x_d) \phi_1(x_1)\phi_1(x_2) \dots \phi_p(x_1)\phi_p(x_2) \dots \phi_p(x_{d-1}) \phi_p(x_d) \mid \phi_1(x_1) \phi_1(x_2) \\
 & \phi_1(x_3) \dots \phi_p(x_1) \phi_p(x_2) \phi_p(x_3) \dots \phi_p(x_1) \phi_p(x_2) \phi_p(x_4) \dots \phi_p(x_{d-2}) \phi_p(x_{d-1}) \phi_p(x_d) \\
 & \phi_1(x_1)\phi_1(x_2) \phi_1(x_1)\phi_1(x_3) \phi_1(x_2)\phi_1(x_3) \dots \phi_p(x_{d-2})\phi_p(x_{d-1}) \phi_p(x_{d-2})\phi_p(x_d) \phi_p(x_{d-1})\phi_p(x_d) \\
 & \phi_1(x_1)\phi_1(x_2)\phi_1(x_3) \dots \phi_p(x_1)\phi_p(x_2)\phi_p(x_3) \dots \phi_p(x_{d-2})\phi_p(x_{d-1})\phi_p(x_d)]^\top
 \end{aligned} \tag{28}$$

and $l \times 1$ coefficient vector with the corresponding multiples, α , β , and γ , of each component function

$$\begin{aligned}
 \mathbf{c} = & \left[\alpha_1^{(1)1} \dots \alpha_1^{(2)2} \dots \alpha_p^{(d)d} \mid \alpha_1^{(12)1} \alpha_1^{(12)2} \dots \alpha_p^{(12)1} \alpha_p^{(12)2} \dots \alpha_p^{(13)1} \alpha_p^{(13)3} \right. \\
 & \dots \alpha_p^{(d-1,d)d-1} \alpha_p^{(d-1,d)d} \beta_{11}^{(12)12} \dots \beta_{pp}^{(12)12} \dots \beta_{pp}^{(d-1,d)d-1,d} \mid \alpha_1^{(123)1} \alpha_1^{(123)2} \\
 & \alpha_1^{(123)3} \dots \alpha_p^{(123)1} \alpha_p^{(123)2} \alpha_p^{(123)3} \dots \alpha_p^{(124)1} \alpha_p^{(124)2} \alpha_p^{(124)4} \dots \\
 & \alpha_p^{(d-2,d-1,d)d-2} \alpha_p^{(d-2,d-1,d)d-1} \alpha_p^{(d-2,d-1,d)d} \beta_{11}^{(123)12} \beta_{11}^{(123)13} \beta_{11}^{(123)23} \\
 & \dots \beta_{pp}^{(d-2,d-1,d)d-2,d-1} \beta_{pp}^{(d-2,d-1,d)d-2,d} \beta_{pp}^{(d-2,d-1,d)d-2,d-1,d} \\
 & \left. \gamma_{111}^{(123)123} \dots \gamma_{ppp}^{(123)123} \dots \gamma_{ppp}^{(d-2,d-1,d)d-2,d-1,d} \right]^\top,
 \end{aligned} \tag{29}$$

where the symbol T denotes transpose. For a collection of N input vectors, $\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)}\}$, and corresponding function values, $\{f(\mathbf{x}^{(1)}), f(\mathbf{x}^{(2)}), \dots, f(\mathbf{x}^{(N)})\}$, we must repeat the vector inner product of Equation 27 a total of N times. This system of linear equations can be cast in matrix form

$$\Phi \mathbf{c} = \mathbf{b}, \tag{30}$$

where the $N \times l$ design matrix, Φ , and $N \times 1$ data vector, \mathbf{b} , have the following entries

$$\Phi = \begin{bmatrix} \phi(\mathbf{x}^{(1)})^\top \\ \phi(\mathbf{x}^{(2)})^\top \\ \vdots \\ \phi(\mathbf{x}^{(N)})^\top \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} f(\mathbf{x}^{(1)}) - f_0 \\ f(\mathbf{x}^{(2)}) - f_0 \\ \vdots \\ f(\mathbf{x}^{(N)}) - f_0 \end{bmatrix}. \quad (31)$$

Thus, the design matrix stores row-wise the design vectors of the collection of \mathbf{x} 's and the data vector equals the right-hand side of Equation 27 for each of these N input vectors.

The first-, second-, and third-order component functions of Equation 24 that enter into the model output decomposition of Equation 25 are linear in their respective coefficients, α , β , and γ . The design matrix, Φ , thus, is comprised of valid basis functions and we can solve directly for the l -vector of unknown coefficients, \mathbf{c} , using linear least squares. Yet, the matrix form of Equation 30 does not preserve hierarchical orthogonality of the component functions. Furthermore, this default formulation does not offer any protection against underdetermined problems due to using either an insufficient number $N < l$ of $(\mathbf{x}, f(\mathbf{x}))$ -data pairs or a rank-deficient design matrix, Φ . Then, the linear system of algebraic Equations in Equation 30 will have infinite solutions, $\hat{\mathbf{c}}_{ls} = (\Phi^\top \Phi)^{-1} \Phi^\top \mathbf{b}$, and the expansion coefficients of the function decomposition of Equation 25 may be ill-defined with poor an/or uncertain estimates of the sensitivity indices. Fortunately, if one looks carefully at the design vector in Equation 28 then one will notice that some of the (polynomial) basis functions appear more than once in $\Phi(\mathbf{x})$. For example, $\phi_1(x_1)$ pops up in the entries of the first-, second-, and third-order component functions (separated by a vertical bar). This suggests that we can simplify the matrix form of Equation 30 to a reduced system of linear Equations which is easier to solve in practice.

Li and Rabitz (2012) introduced a solution method coined Diffeomorphic modulation under observable response preserving homotopy, or D-MORPH regression, which satisfies hierarchical orthogonality of the component functions and simplifies the search for the preferred optimum coefficient vector, $\hat{\mathbf{c}}_{dm}$, among the potentially infinite number of plausible solutions. We provide a brief step-by-step summary of D-MORPH regression and refer interested readers to Li and Rabitz (2010) for a more exhaustive description. In short, the least squares values of the expansion coefficients may be computed as follows

$$\hat{\mathbf{c}}_{ls} = (\Phi^\top \Phi)^\dagger \mathbf{d}, \quad (32)$$

where the $l \times (l - d \cdot p)$ matrix $(\Phi^\top \Phi)^\dagger$ is the generalized pseudo inverse of the $l \times l$ Gramian matrix, $\mathbf{G} = \Phi^\top \Phi$, satisfying all four Moore-Penrose conditions (Golub & Van Loan, 1996; Penrose, 1955), whose redundant rows (the first $d \cdot p$ rows of first-order basis functions) are removed and \mathbf{d} is the $(l - d \cdot p) \times 1$ vector $\Phi^\top \mathbf{b}$ without the first $d \cdot p$ rows. Whether unique or not, the least squares solution of Equation 32 does not enforce hierarchical orthogonality of the first-, second-, and third-order component functions. Therefore, Li and Rabitz (2010) introduce a square constrained matrix, \mathbf{B} , which consists of the inner products of the orthonormalized polynomial basis functions of Equation 28 so as to satisfy the relaxed vanishing condition of Equation 22

$$\int_0^1 w_i(x_i) f_0 f_i(x_i) dx_i = 0 \quad (33a)$$

$$\int_0^1 w_{ij}(x_i, x_j) f_i(x_i) f_{ij}(x_i, x_j) dx_i dx_j = 0 \quad (33b)$$

$$\int_0^1 w_{ijk}(x_i, x_j, x_k) f_{ij}(x_i, x_j) f_{ijk}(x_i, x_j, x_k) dx_i dx_j dx_k = 0. \quad (33c)$$

The expansion coefficients, $\hat{\mathbf{c}}_{dm}$, derived from D-MORPH regression are now written as a linear combination of their values, $\hat{\mathbf{c}}_{ls}$, obtained from least-squares regression

$$\hat{\mathbf{c}}_{dm} = \mathbf{V}_{l-r} (\mathbf{U}_{l-r}^\top \mathbf{V}_{l-r}) \mathbf{U}_{l-r}^\top \hat{\mathbf{c}}_{ls}, \quad (34)$$

where \mathbf{U}_{l-r} and \mathbf{V}_{l-r} equal the last $l - r$ columns of the $l \times l$ matrices \mathbf{U} and \mathbf{V} determined from singular value decomposition (SVD) of the product of a $l \times l$ projection matrix $\mathbf{P} = (\mathbf{I}_l - \mathbf{G})$ and the $l \times l$ constrained matrix \mathbf{B}

Algorithm 1. Main steps of HDMR_{ext}

Input: A collection, $\mathbf{X} = \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}\}$, of N samples of $\mathbf{x} = \{x_1, \dots, x_d\}$ and corresponding function values, $\mathbf{y} = \{y^{(1)}, \dots, y^{(N)}\}$, where $y^{(t)} = f(\mathbf{x}^{(t)})$; $t = (1, \dots, N)$
The order, $\delta \in \{1, 2, 3\}$, of the HDMR_{ext} expansion in Equation (25)
Output: (S_u^a, S_u^b, S_u) for any subset u of one up to δ entries of the superset $U = \{1, \dots, d\}$

- 1: **begin**
- 2: Map the samples of \mathbf{X} to the d -dimensional unit hypercube, \mathbb{K}^d
- 3: Compute a , b and c of polynomials of Equation (23) using Gram-Schmidt orthonormalization
- 4: Determine the $N \times l$ design matrix, Φ , of Equation (31)
- 5: Determine the $N \times 1$ data vector, \mathbf{b} , in Equation (31)
- 6: Solve for the expansion coefficients, $\hat{\mathbf{c}}_{\text{dm}}$, using D-MORPH regression
- 7: Evaluate the component functions in Equation (24) using $\hat{\mathbf{c}}_{\text{dm}}$ and the samples of \mathbf{X} in \mathbb{K}^d
- 8: **for** $u \subseteq \{1, \dots, d\}$; $1 \leq |u| \leq \delta$ **do**
- 9: Compute S_u^a , S_u^b and S_u using Equations (14) and (15)
- 10: **end for**
- 11: **return** $(S_u^a, S_u^b, S_u) \forall u \subseteq \{1, \dots, d\}; 1 \leq |u| \leq \delta$
- 12: **end**

$$\mathbf{PB} = \mathbf{U}\Sigma\mathbf{V}^T, \tag{35}$$

where \mathbf{I}_l signifies the $l \times l$ identity matrix and r denotes the number of nonzero singular values. Upon determination of the D-MORPH values of the expansion coefficients, $\hat{\mathbf{c}}_{\text{dm}}$, of the component functions, we resort to Equations 14 and 15 for the structural, S_u^a , correlative, S_u^b , and total, S_u , sensitivity indices of each input variable and combinations thereof.

Hierarchical orthogonality of the component functions imposes a strong constraint on the expansion coefficients. This does not, however, relax the requirements of a sufficiently large sample N of $(\mathbf{x}, f(\mathbf{x}))$ -data pairs. To guarantee robust estimates of the sensitivity indices with relatively small uncertainty, the number of input-output samples N must be substantially larger than the number of expansion coefficients l . In our case studies, the ratio N/l exceeds 10. Pending a sufficient order δ of the model decomposition, the use of orthonormalized polynomial functions with extended bases in conjunction with D-MORPH regression guarantees an exact variance decomposition of the model output.

The default implementation of HDMR_{ext} assumes the third-order approximation of Equation 25. If so desired, the user can specify a lower order expansion of the model output. For example, for $\delta = 2$, the HDMR_{ext} model decomposition would use first- and/or second-order component functions only. This will involve substantially fewer expansion coefficients, l , and is particularly attractive for high-dimensional input spaces, \mathbb{K}^d . Of course, care should always be exercised that the HDMR_{ext} expansion of the model output is sufficiently accurate. Unless stated differently, we further assume that $p = 3$, thus, use the orthonormalized polynomial functions of first-, second- and third-degree of Equation 23. This amounts to the component functions of Equation 24.

Algorithm 1 and Figure S2 in Supporting Information S1 summarize the different steps of the HDMR_{ext} method. The user must specify the maximum order, $1 \leq \delta \leq 3$; $\delta \in \mathbb{N}_+$, of the HDMR_{ext} component functions and supply a collection, $\mathbf{X} = \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(N)}\}$, of N samples of $\mathbf{x} = \{x_1, \dots, x_d\}$ with corresponding function values, $\mathbf{y} = \{y^{(1)}, \dots, y^{(N)}\}$. These d -vectors of input variables may be drawn from an arbitrary joint probability distribution, $p(\mathbf{x})$, specified on the prior domain χ . In the context of probabilistic SA the samples of \mathbf{X} amount to realizations of the posterior distribution, $p(\mathbf{x}|\tilde{\mathbf{y}})$. Prior to computation of the design matrix, Φ , the samples of \mathbf{X} are mapped to the unit hypercube, \mathbb{K}^d .

4. Case Studies

We illustrate the HDMR and HDMR_{ext} methods by application to variance-based SA of the parameters of four different models with increasing complexity. The first two studies serve as illustration of both methods and benchmark the performance of HDMR and HDMR_{ext} against known sensitivity indices of the parameters. The last two studies present the application of variance/covariance-based SA to hydrologic modeling. As a reminder, in each study, the input variables are normalized/rescaled to the unit interval of the cubic B-splines and orthogonal polynomial functions prior to application of HDMR and HDMR_{ext} . The structural and correlative sensitivity indices, S_i^a and S_i^b , are color coded in our figures using “apple green” and “blue” as mnemonic of the superscripts

“a” and “b”, respectively. Table S1 in Supporting Information S1 specifies the values of d and N for each case study along with our settings for the algorithmic variables of HDMR and HDMR_{ext}. These tabulated values confirm that (a) the ratio N/d is at least equal to 7 and (b) the number of expansion coefficients of HDMR exceeds that of HDMR_{ext}. As this paper has a methodological focus we chose rather common hydrologic case studies of variance-based SA. This should make it easier to digest the results. As these cases involve the use of relatively simple functions and models, computational efficiency is not a key consideration.

4.1. A Linear Function

We follow Li et al. (2010) and consider a simple additive function

$$y = f(\mathbf{x}) = \sum_{i=1}^4 x_i + e, \quad (36)$$

where the $d = 4$ input variables, x_1 , x_2 , x_3 , and x_4 are drawn jointly from a multivariate normal distribution, $\mathbf{x} \sim \mathcal{N}_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with unit mean $\boldsymbol{\mu} = [1 \ 1 \ 1 \ 1]^T$ and 4×4 covariance matrix, $\boldsymbol{\Sigma}$, and the noise term, $e \sim \mathcal{N}(0, \sigma_e^2)$ with $\sigma_e^2 \ll \text{Var}[y]$. We evaluate the function for two different collections of five-thousand samples, \mathbf{X}_1 and \mathbf{X}_2 , that differ in their covariance matrix

$$\boldsymbol{\Sigma}_1 = \begin{bmatrix} 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix} \quad \boldsymbol{\Sigma}_2 = \begin{bmatrix} 1.0 & 0.6 & 0.2 & 0.0 \\ 0.6 & 1.0 & 0.2 & 0.0 \\ 0.2 & 0.2 & 1.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 \end{bmatrix}, \quad (37)$$

As the off-diagonal entries of $\boldsymbol{\Sigma}_1$ are zero, the entries of x in the first collection of N samples, \mathbf{X}_1 , are uncorrelated. The samples of collection \mathbf{X}_2 , on the contrary, are dependent with Pearson correlation coefficient of 0.6 (x_1 and x_2), 0.2 (x_1 and x_3) and 0.2 (x_2 and x_3), respectively. We implement the bootstrap method (Efron & Tibshirani, 1994; Storlie et al., 2009) and execute HDMR and HDMR_{ext} a total of $M = 100$ different times with each trial using a different selection of $N = 3,000$ samples (drawn without replacement) from the original collections, \mathbf{X}_1 and \mathbf{X}_2 , of five-thousand parameter vectors. The linear function in Equation 36 does not have interaction terms, thus, a maximum order, $\delta = 1$, will suffice for HDMR and HDMR_{ext}.

Table 1 reports mean values of the structural, S_i^a , correlative, S_i^b , and total, S_i^t , sensitivity indices of the four input variables (parameters) of the linear function in Equation 36 for the collection of independent input vectors, \mathbf{X}_1 , using HDMR (left-side) and HDMR_{ext} (right-side). The values between parenthesis document the 99% confidence intervals of the bootstrap trials. The results of HDMR and HDMR_{ext} are almost indistinguishable, as is to be expected when the input variables are independent. The structural sensitivity indices confirm that all four input variables have an equal contribution to the model output, y . The correlative sensitivity indices of about zero testify to the independence of the input variables. The 99% confidence intervals of the sensitivity indices appear very small as a result of using a relatively large number of samples N in each bootstrap trial relative to the number of unknown coefficients. Thus, the component functions are very well defined in each separate bootstrap trial.

We now move on to the collection \mathbf{X}_2 of dependent input variables and present in Table 2 mean values and associated 99% confidence intervals (between parenthesis) of the structural, correlative and total sensitivity indices of HDMR and HDMR_{ext} derived from the bootstrap method using hundred different trials. The results of HDMR and HDMR_{ext} are again in close agreement but the presence of correlation among the input variables has a noticeable impact on the sensitivity indices of HDMR and HDMR_{ext}. The direct contributions of x_1 , x_2 , x_3 , and x_4 to the simulated output of the linear model has lowered somewhat with structural sensitivity indices of about 0.17. The correlative sensitivity indices of all input variables but x_4 have changed from zero to values of about 0.13 for x_1 and x_2 and 0.06 for x_3 . The magnitude of these values is in agreement with the covariance matrix, $\boldsymbol{\Sigma}_2$, of the samples. Indeed, if we take the sum of the covariances in each row then we observe a 2:1 ratio in the total covariance of x_1 (or x_2) and x_3 . This simple back of the envelope calculation confirms that the correlative contribution of x_1 and x_2 should be about twice that of x_3 . Per the covariance matrix, $\boldsymbol{\Sigma}_2$, the fourth input variable, x_4 , is independent of the other three inputs. Hence, its correlative sensitivity index should be zero. Note that Sobol's method would only provide estimates of the structural sensitivity.

Table 1
Structural, Correlative, and Total Sensitivity Indices of High-Dimensional Model Representation (HDMR) and HDMR_{ext} for a Linear Model With Independent Input Variables

Term	HDMR			HDMR _{ext}		
	S_i^a	S_i^b	S_i	S_i^a	S_i^b	S_i
x_1	0.251 (±0.010)	-0.001 (±0.011)	0.251 (±0.011)	0.244 (±0.010)	-0.002 (±0.009)	0.242 (±0.010)
x_2	0.252 (±0.010)	0.001 (±0.009)	0.253 (±0.009)	0.251 (±0.010)	-0.000 (±0.009)	0.251 (±0.009)
x_3	0.248 (±0.011)	0.000 (±0.010)	0.248 (±0.011)	0.251 (±0.010)	0.003 (±0.009)	0.254 (±0.009)
x_4	0.246 (±0.009)	0.002 (±0.011)	0.248 (±0.010)	0.247 (±0.009)	0.005 (±0.010)	0.253 (±0.009)
Sum	0.998 (±0.029)	0.002 (±0.029)	1.000 (±0.000)	0.994 (±0.028)	0.006 (±0.028)	1.000 (±0.000)

We conclude that HDMR and HDMR_{ext} yield similar results for a simple additive function with dependent or independent input variables. The values of the structural, correlative and total sensitivity indices are in agreement with prior expectations. Given the nature of the linear function, in HDMR and HDMR_{ext} we used a first-order approximation only of the simulated output. One can expand the emulator with second- and third-order component functions, but this would not change the results as these terms should/will not contribute to the model output. One reviewer alerted to the smaller confidence intervals of S_i^a , S_i^b and S_i for the dependent input variables. Albeit in agreement with Li et al. (2010), this result may not be intuitive. The covariance matrix Σ_2 enlarges the variance of the model output, which, in turn reduces the uncertainty of the structural, correlative and total sensitivity indices.

4.2. The Ishigami Function

As our second case study, we consider the somewhat more complex Ishigami function. This function is a recurring test case for SA methods due to its non-linear and non-monotonic behavior and presence of interaction effects between the input variables. The Ishigami function is given by Ishigami and Homma (1990)

$$y = f(\mathbf{x}) = \sin(x_1) + a\sin^2(x_2) + bx_3^4\sin(x_1), \tag{38}$$

where $\mathbf{x} = \{x_1, x_2, x_3\} \in [-\pi, \pi]^3$ signify the input variables and a and b are constants set to 7 and 0.1, respectively (Marrel et al., 2009). The non-monotonic behavior of the Ishigami function warrants use of a larger number of knots and polynomial degree for the component functions of HDMR and HDMR_{ext}, respectively. Based on convergence analysis (see Figure S3 in Supporting Information S1) of the sample variance s_e^2 of the functional decomposition of Equations 19 and 25, we use $m = 5$ Bézier curves (HDMR) and a sixth-order polynomial $p = 6$ (HDMR_{ext}) instead (e.g., Table S1 in Supporting Information S1).

In our first experiment, we draw $N = 8,000$ realizations, $\mathbf{X} = \{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(N)}\}$ from the trivariate uniform prior distribution, $\mathbf{x} \sim \mathcal{U}_3(-\pi, \pi)$ between $-\pi$ and π using Latin hypercube sampling. If computational efficiency is of key importance, then one can use Sobol' (1967) sequences instead as method for quasi-random sampling (Kucherenko et al., 2011, 2015; Zuniga et al., 2013). For each sample, we evaluate the Ishigami function and admit the collection of $N = 8,000$ different (\mathbf{x}, y) -data pairs to HDMR and HDMR_{ext}. We use $\delta = 2$, thus, discard the third-order component functions in our function approximation with HDMR and HDMR_{ext}.

Table 2
Structural, Correlative, and Total Sensitivity Indices of High-Dimensional Model Representation and HDMR_{ext} for a Linear Model With Dependent Input Variables

Term	HDMR			HDMR _{ext}		
	S_i^a	S_i^b	S_i	S_i^a	S_i^b	S_i
x_1	0.164 (±0.006)	0.135 (±0.004)	0.299 (±0.007)	0.171 (±0.006)	0.136 (±0.004)	0.307 (±0.006)
x_2	0.169 (±0.004)	0.132 (±0.004)	0.301 (±0.006)	0.167 (±0.005)	0.132 (±0.004)	0.300 (±0.006)
x_3	0.172 (±0.007)	0.064 (±0.007)	0.235 (±0.008)	0.169 (±0.007)	0.067 (±0.006)	0.236 (±0.007)
x_4	0.165 (±0.007)	-0.004 (±0.009)	0.161 (±0.009)	0.165 (±0.006)	-0.009 (±0.010)	0.157 (±0.008)
Sum	0.671 (±0.014)	0.329 (±0.014)	1.000 (±0.000)	0.673 (±0.017)	0.327 (±0.017)	1.000 (±0.000)

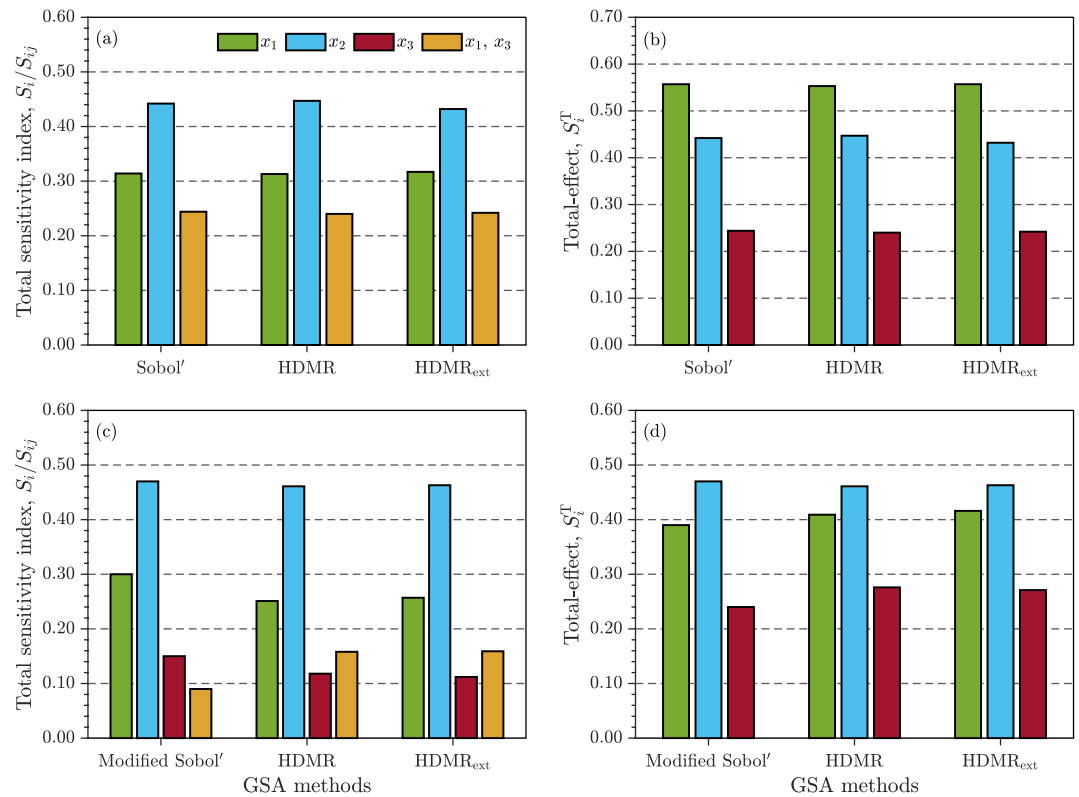


Figure 1. Comparison of sensitivity indices (total indices of individual terms or combinations thereof, S_i/S_{ij} and total-effect, S_i^T) of Ishigami function derived from different GSA methods, where we consider uncorrelated inputs for graph (a) and (b) while correlated inputs ($\rho(x_1, x_3) = 0.5$) for graph (c) and (d). Note that in graph (a), the third entry of the bar plot corresponds to the interaction between x_1 and x_3 because there is not individual contribution of x_3 to the model (i.e., $S_3 = 0$).

Figure 1 presents bar charts of the (a) total sensitivity index, S_i/S_{ij} , of Equation 15 and (b) total-effect, S_i^T of Equation 5, of the three input variables of the Ishigami function derived from Sobol's method, HDMR and HDMR_{ext} using the collection of $N = 8,000$ independent samples from the uniform prior distribution. The left bar chart also display the sensitivity index of x_1 and x_3 combined (in yellow). The total sensitivity indices and total effects of HDMR and HDMR_{ext} are not only in excellent mutual agreement (as expected) but are also an excellent approximation to the analytic sensitivity estimates of the Ishigami function articulated by Sobol's method (Homma & Saltelli, 1996). The third input variable, x_3 , does not have a first-order contribution to the output of the Ishigami function. Hence, its total sensitivity index, S_3 , is zero in Figure 1a. This is also true for the total sensitivity indices, S_{12} , S_{23} , and S_{123} , which do not contribute to the model output as is evident from the mathematical definition of the Ishigami function. The only interaction that counts originates from x_1 and x_3 . We do not tabulate the uncertainty estimates of the sensitivity indices. The 99% confidence intervals of the total sensitivity indices of HDMR and HDMR_{ext} are very small and on the order of ± 0.01 for bootstrapping with 4,000 input-output samples.

We now turn our attention to the collection \mathbf{X}_2 of correlated input vectors. The bottom panel of Figure 1 presents the same indices as in the top panel for (modified) Sobol's method, HDMR and HDMR_{ext} but using a collection of $N = 8,000$ samples of \mathbf{x} with correlation coefficient of 0.5 between the values of x_1 and x_3 . The modified Sobol method of Kucherenko et al. (2012) is a generalization of the variance-based sensitivity indices of Sobol to dependent input variables. Notice again the excellent agreement between the three different methods. The correlation between x_1 and x_3 increases the first-order contribution of x_3 from zero to about 0.12. This in turn reduces the first-order effect of x_1 and the second-order effect of (x_1, x_3) . This changes the ranking of the total-effects of the three input variables from $S_1^T > S_2^T > S_3^T$ with independent inputs to $S_2^T > S_1^T > S_3^T$ with correlation among x_1 and x_3 . Thus, correlation between the input factors of a model can profoundly change the relative importance of each input variable. The 99% confidence intervals of the HDMR/HDMR_{ext} sensitivity indices have only marginally increased compared to the case with independent variables to an average of about ± 0.012 for S_1, S_2, S_3, S_{13} .

This second study confirms the ability and robustness of HDMR and HDMR_{ext} to decompose the variance of a simple nonlinear function with interaction between some of its input variables. From hereon, we focus our attention on more practical problems using examples from hydrologic modeling. In these studies we use relatively large values of the ratio N/l and, consequently, the 99% confidence intervals of the sensitivity indices will be small. As a result, we do not report the bootstrap uncertainty estimates. Furthermore, as our interest is in a rigorous variance-based decomposition of the input-output relationship at first-, second- and higher orders, we do not further consider the modified Sobol method of Kucherenko et al. (2012). In fact, in reference to this and the modified FAST method of Xu (2013), Song et al. (2015) even goes on to say (p. 751) that there “...have been very few successful applications in hydrological modeling.”

4.3. The Water Retention Function

4.3.1. Model Description

The third case study considers a common problem in vadose zone hydrology and involves the characterization of the hydraulic properties of variably saturated soils. This study serves to demonstrate the ability of HDMR and HDMR_{ext} to guide experimental design and help determine measurement collection. We use the capillary pressure-saturation function of van Genuchten (1980) which describes the relationship between the volumetric soil moisture content, θ [$\text{L}^3 \text{L}^{-3}$], and soil water pressure head, h [L], as follows

$$\theta(h) = \theta_r + (\theta_s - \theta_r) \left[1 + (\alpha_{\text{vg}} |h|)^{n_{\text{vg}}} \right]^{1/n_{\text{vg}} - 1}, \quad (39)$$

where θ_s and θ_r [$\text{L}^3 \text{L}^{-3}$] denote the saturated and residual moisture contents, respectively, α_{vg} [L^{-1}] is an estimate of the reciprocal of the air-entry value and n_{vg} is a dimensionless scalar. The soil hydraulic parameters are considered input factors in our present analysis, thus, $\mathbf{x} = \{\theta_r, \theta_s, \alpha_{\text{vg}}, n_{\text{vg}}\}$. To investigate the relative importance of each hydraulic parameter to the model output, θ , we generate a vector of two-hundred logarithmically equally spaced points of the soil water pressure head between decades -10^6 and -10^{-1} cm. Next, we create the $N \times d$ matrices, \mathbf{X}_1 and \mathbf{X}_2 , by drawing $N = 5,000$ parameter vectors from a multivariate normal distribution, $\mathbf{x} \sim \mathcal{N}_4(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with mean, $\boldsymbol{\mu}$, $d \times 1$ vector of standard deviations, $\boldsymbol{\sigma}$, and $d \times d$ covariance matrices, $\boldsymbol{\Sigma}_1 = \mathbf{R}_1(\boldsymbol{\sigma}\boldsymbol{\sigma}^\top)$ and $\boldsymbol{\Sigma}_2 = \mathbf{R}_2(\boldsymbol{\sigma}\boldsymbol{\sigma}^\top)$, specified in Table S2 in Supporting Information S1. These listed statistics are taken from a study by Gomes et al. (2017) on the stability of a variably-saturated slope in Rio de Janeiro, Brazil. Thus, the input factors of \mathbf{X}_1 are independent, whereas correlation is present between the input variables of \mathbf{X}_2 . We consider a third-order approximation of the model output, $\delta = 3$, and execute the HDMR and HDMR_{ext} methods for each discretized soil water pressure head, h . This produces two-hundred values of the structural, correlative and total sensitivity indices, S_i^a , S_i^b , and S_i^T , respectively, for individual input variables and combinations thereof. We separately also compute the total-effect, S_i^T , of each input factor.

After a careful inspection of the results of both SA methods, we decided to discard the results of HDMR and focus exclusively on HDMR_{ext} in this case study. The sensitivity indices derived from HDMR reveal a problem with the variance decomposition of Equation 39 in the face of correlated input variables (see Figure S4 in Supporting Information S1). The total sensitivity indices of the n_{123} component functions of HDMR do not sum to one. This cannot be resolved by using a larger sample size N but is simply a consequence of backfitting and failure of HDMR to satisfy hierarchical orthogonality of the component functions. A larger number of control points (or Bézier curves, m) may sometimes help to relieve this problem but this is not a permanent remedy. The deviation of the sum of the total sensitivity indices from unity is small at volumetric moisture contents close to residual saturation (left-side) but increases substantially at some of the soil water pressure heads in the wet range of the WRF. This inexact variance/covariance decomposition will corrupt and/or bias the structural, correlative and total sensitivity indices of the input factors. The third-order function approximation in Equation 25 does not suffer this problem. The total sensitivity indices computed by HDMR_{ext} consistently sum up to one for all soil water pressure heads. This is a testament to the hierarchical orthogonality of the component functions.

4.3.2. Uncorrelated Input Variables

Figure 2 displays the structural, S_i^a , and correlative, S_i^b , sensitivity indices and total-effect, S_i^T , of (a) θ_r , (b) θ_s , (c) α_{vg} , and (d) n_{vg} as function of the soil water pressure head, h , using dashed apple green, dashed blue and solid orange lines, respectively. For completeness, the solid black lines depict the absolute values of the first-order partial derivatives, $\partial\theta/\partial x_i$, for all $i = (1, \dots, 4)$, of the discretized model input, h_j , at $\mathbf{x} = \boldsymbol{\mu}$, where $j = (1, \dots, 200)$. These

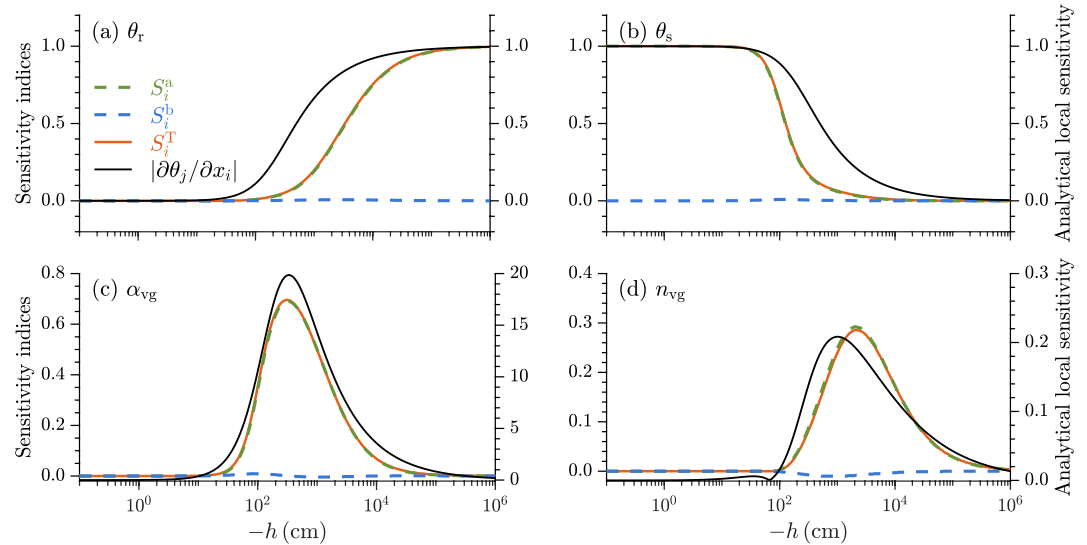


Figure 2. Uncorrelated soil hydraulic parameters: Structural, S_i^a , and correlative, S_i^b , sensitivity indices and total-effect, S_i^T , of (a) θ_r , (b) θ_s , (c) α_{vg} , and (d) n_{vg} as function of the soil water pressure head. The solid black lines display the first-order partial derivatives, $\partial\theta_j/\partial x_i$, of the soil moisture content, θ , with respect to each individual input factor.

analytic values portray the instantaneous rate of change of the soil water pressure head and quantify the local sensitivity of each hydraulic parameter. Note that the values of $\partial\theta_j/\partial x_i$ make up the entries of the well-known Jacobian (or sensitivity) matrix at $\mathbf{x} = \boldsymbol{\mu}$. The structural sensitivity indices exhibit a (a, b) sigmoidal and (c, d) bell-shaped relationship with the soil water pressure head. As expected, the correlative sensitivity indices are constant at zero. This is a result of using uncorrelated input factors in the variance/covariance decomposition of Equation 25.

The graphs of the structural sensitivity indices bear a close resemblance to the local sensitivities (solid black lines) of the hydraulic parameters. This is not an uncommon finding and corroborated by the theoretical treatise of Kucherenko and Song (2016) which links global derivative-based measures of sensitivity to Sobol' total-effect index S_i^T in Equation 5. For independent input variables, derivative-based sensitivities can offer a valuable CPU-efficient alternative to variance-based methods for computationally costly and/or highly-parameterized nonlinear dynamic models (Kiparissides et al., 2009). But we may not compare the values of the two sensitivity measures as they have different units and interpretations. The variance-based sensitivity indices of HDMR_{ext} are normalized, thus, dimensionless, and characterize the mean sensitivity of the input factors in the neighborhood of their mean values. The first-order partial derivatives have units of cm or $\text{cm}^3 \text{cm}^{-3}$ and quantify only the local sensitivity of the model output, θ , with respect to each input factor. These local sensitivities are intimately related to the Fisher information matrix and serve an important purpose in experimental design. For functions that are nonlinear in their parameters, the global variance-based indices of HDMR_{ext} will not match perfectly with the local sensitivity estimates of θ_r , θ_s , α_{vg} , and n_{vg} unless we confine the sample of input vectors of HDMR_{ext} to a small space immediately surrounding their mean values of Table S2 in Supporting Information S1.

The moisture content of the soil exerts a large control on the structural sensitivity of the hydraulic parameters. The most informative (θ, h) -data pairs for θ_r , θ_s , α_{vg} , and n_{vg} are beyond the soil's wilting point ($h \in [-\infty, -1.5 \times 10^4]$ cm), close to saturation ($h \in [-10, 0]$ cm), near the air-entry value ($h \in [-300, -200]$ cm) and around the inflection point ($h \in [-1, 200, -1, 000]$ cm) of the WRF, respectively. This confirms earlier findings of Vrugt et al. (2002) and Vrugt and ter Braak (2011). Note that the listed ranges of the soil water pressure head, h , are not particularly meaningful as they are soil dependent. The total-effect of θ_r , θ_s , α_{vg} , and n_{vg} is indistinguishable from the structural sensitivity of each hydraulic parameter. Thus, the second- and third-order component functions of the HDMR_{ext} emulator must have a negligible contribution to the model output, θ . This is not uncommon for simple functions and/or uncorrelated input factors.

4.3.3. Correlated Input Variables

The use of independent input variables does not do justice to the strong correlations found among the hydraulic parameters in laboratory and field experiments. Based on our results for the Ishigami function, we can only

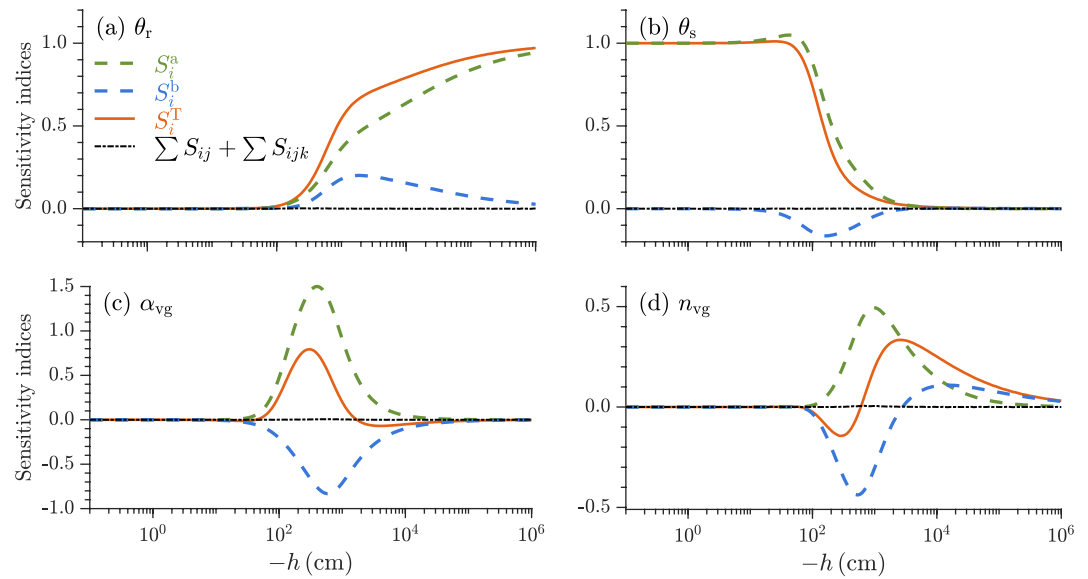


Figure 3. Correlated soil hydraulic parameters: The structural, S_i^a , and correlative, S_i^b , sensitivity indices and total-effect, S_i^T , of (a) θ_r , (b) θ_s , (c) α_{vg} , and (d) n_{vg} as function of the soil water pressure head. The dashed-dotted black line plots the sum of total sensitivity indices of the second and third-order interactions (component functions).

posit that this correlation will impact the structural sensitivities of the WRF parameters, θ_r , θ_s , α_{vg} , and n_{vg} , and govern their total effects. This may have implications for experimental design and the precise location of the most informative (θ, h) -data points along the WRF.

Figure 3 presents the structural sensitivity, S_i^a , correlative sensitivity, S_i^b , and total-effect, S_i^T , of (a) θ_r , (b) θ_s , (c) α_{vg} , and (d) n_{vg} as function of the soil water pressure head, h , using dashed apple green, dashed blue, and solid orange lines, respectively. We also display the higher-order effect, $\sum S_{ij} + \sum S_{ijk}$, using dash-dotted black lines. The correlation among the four soil hydraulic parameters manifests itself in non-zero values of the correlative sensitivity indices, S_i^b . As a result of this, the structural sensitivities S_i^a of θ_s , θ_r , α_{vg} , and n_{vg} do not equal their total-effects S_i^T . The correlation, however, does not impact the parameters' total-effects across the entire range of soil water pressure heads. Rather, the correlative contribution of the soil hydraulic parameters to the model output, θ , is limited to $h \in [-10^6, -10^2]$ cm. In this interval, the soil's volumetric water content decreases from near saturation to residual moisture content. Outside this range, the correlative contribution of θ_r , θ_s , α_{vg} , and n_{vg} to the model output is near zero. The magnitude of the correlative sensitivity indices is commensurate with the matrix of correlation coefficients, \mathbf{R}_2 , of Table S2 in Supporting Information S1.

The negative values of the correlative sensitivity indices of α_{vg} and n_{vg} are debit to their negative total-effects in the range of $h \in [-10^5, -10^3]$ and $h \in [-600, -70]$ cm, respectively. Negative S_i^T 's are not possible with Sobol's method and, thus, may seem odd at first sight. But covariances among the input variables can suppress variations in model output, hence decrease their total influence on the model output. This result is confirmed by a simple analytic example in Text S1 in Supporting Information S1.

The strong negative correlation, $R_{\alpha_{vg}, n_{vg}} = -0.86$, between α_{vg} and n_{vg} is not only responsible for the negative total-effects of α_{vg} and n_{vg} over the documented range of soil water pressure heads but, in turn, also increases the magnitude of the structural sensitivities of α_{vg} and n_{vg} . Nevertheless, the graphs of Figures 2 and 3 are in sufficient agreement not to alter much our earlier conclusions with respect to experimental design, where one should pay attention to only the sensitive domain of correlated parameters for the appropriate selection of informative $\theta(h)$ data pairs.

The secondary and third-order component functions again have a negligible contribution to the total effect of the hydraulic parameters. A first-order approximation suffices to accurately describe the $\theta(h)$ relationship of the WRF of van Genuchten (1980) in the presence of correlated and uncorrelated hydraulic parameters. This supports earlier conclusions of Rabitz and Aliş (1999) that most physical systems are adequately characterized by only low-order input-output relationships. Altogether, we conclude that the correlation between the soil hydraulic parameters of the WRF does not fundamentally change our earlier findings with independent input factors.

Thus far we have focused our attention on the structural and correlative sensitivity indices and total-effect of the hydraulic parameters. But the model output decomposition of HDMR_{ext} may serve another important role, and that is as an emulator of the input-output relationship. Figure S5 in Supporting Information S1 compares the third-order approximation of Equation 25 with expansion coefficients derived from D-MORPH regression against the WRF of van Genuchten (1980) using hydraulic parameter values of 12 USDA soils from Carsel and Parrish (1988).

The HDMR_{ext} emulators provide an excellent description of the WRF's of the different soil types. The surrogate WRF's (solid lines) pass exactly through the one-hundred $\theta(h)$ -data pairs of each soil. This inspires confidence in the ability of HDMR_{ext} to provide an accurate assessment of the structural and/or correlative sensitivity indices of the soil hydraulic parameters and combinations thereof. Results demonstrate that the function decomposition of Equation 25 may serve as surrogate of the WRF of Equation 39.

4.4. A Conceptual Rainfall-Runoff Model

4.4.1. Model Description

As fourth and last case study, we turn our attention to hmodel, a parsimonious conceptual watershed model originally developed by Schoups et al. (2010). The hmodel transforms rainfall into runoff at the watershed outlet using an interception, unsaturated zone, fast and slow flow reservoir, respectively, which simulate interception, throughfall, evaporation, surface runoff, percolation, fast streamflow and baseflow (see Figure S6 in Supporting Information S1). In this study, we focus our attention on the parameters of the hmodel. Table S3 in Supporting Information S1 presents a description of each parameter and lists the corresponding symbol, unit, lower and upper bounds. These bounds define a $d = 7$ -dimensional (hyper)cube, $\mathbb{H}_0 \subseteq \mathbb{R}^d$, in which the input factors, $\mathbf{x} = \{I_{\text{max}}, R_{\text{u,max}}, Q_{\text{s,max}}, \alpha_e, \alpha_f, K_f, K_s\}$, are allowed to vary freely.

Many different studies have appeared in the hydrologic literature on the quantification and analysis of parameter sensitivity in conceptual watershed models (Borgonovo et al., 2017; Mockler et al., 2016; A. Wang, Pianosi, et al., 2020; A. Wang & Solomatine, 2019; Zelelew & Alfredsen, 2013). With a few exceptions, these published studies almost always use some form of uniform sampling within the hypercube, \mathbb{H}_0 , to satisfy the requirements of the d -way ANOVA decomposition in Sobol's method and quantify the first-, and/or second-, third- and higher-order parameter sensitivities and/or interactions. We are concerned with the impact of parameter correlation on variance-based estimates of parameter sensitivity and how the structural, correlative and total sensitivity indices depend on the simulated discharge and state (e.g., wetness) of the watershed. As we will show next, the HDMR_{ext} methodology admits the characterization of parameter sensitivity throughout the posterior parameter distribution.

4.4.2. Model Training and Description of Multivariate Parameter Distribution

We estimate the posterior distribution of the hmodel parameters by application to a 9-year record (1 October 1999–30 September 2008) of daily discharge measurements from the Leaf River near Collins, MS (USGS 02472000). This medium-sized watershed of about 1,990 km² exhibits a strong winter regime according to the functional classification of Brunner et al. (2020) and has been studied extensively in the hydrologic literature. Daily time series of streamflow, precipitation, and surface temperature are taken from the CAMELS data set (Newman et al., 2014). We impose a uniform prior distribution over the ranges specified in Table S3 in Supporting Information S1 and use the generalized likelihood function of Schoups and Vrugt (2010) to infer the posterior distribution of the hmodel parameters using Markov chain Monte Carlo simulation with the DREAM_(ZS) algorithm (Vrugt, 2016). The transition kernel of DREAM creates multiple different sequences (chains) of parameter vectors that are stationary and ergodic and have as a joint distribution the posterior target distribution (ter Braak, 2006; Vrugt, ter Braak, et al., 2009). Convergence of the joint chains to a stationary distribution is monitored using single chain and multi-chain diagnostics, among the univariate and multivariate scale-reduction factors of Gelman and Rubin (1992) and Brooks and Gelman (1998), respectively.

Figure 4 presents a scatter plot matrix of the $N = 10,000$ posterior realizations of the hmodel parameters sampled with the DREAM_(ZS) algorithm. The diagonal entries correspond to the parameters' marginal distributions, whereas the off-diagonal graphs present bivariate scatter plots of the different parameter pairs. The Maximum A Posteriori (MAP) hmodel parameter values are separately indicated in each graph with a red cross. The dotted rectangles make up a closed hypercube, $\mathbb{H}_1 \subseteq \mathbb{H}_0$, interior to the prior hypercube, \mathbb{H}_0 . This posterior hypercube will be populated with independent samples so as to compare the sensitivity indices of the posterior realizations of the DREAM_(ZS) algorithm. The hmodel parameters appear well defined by the measured daily discharge data. The histograms of

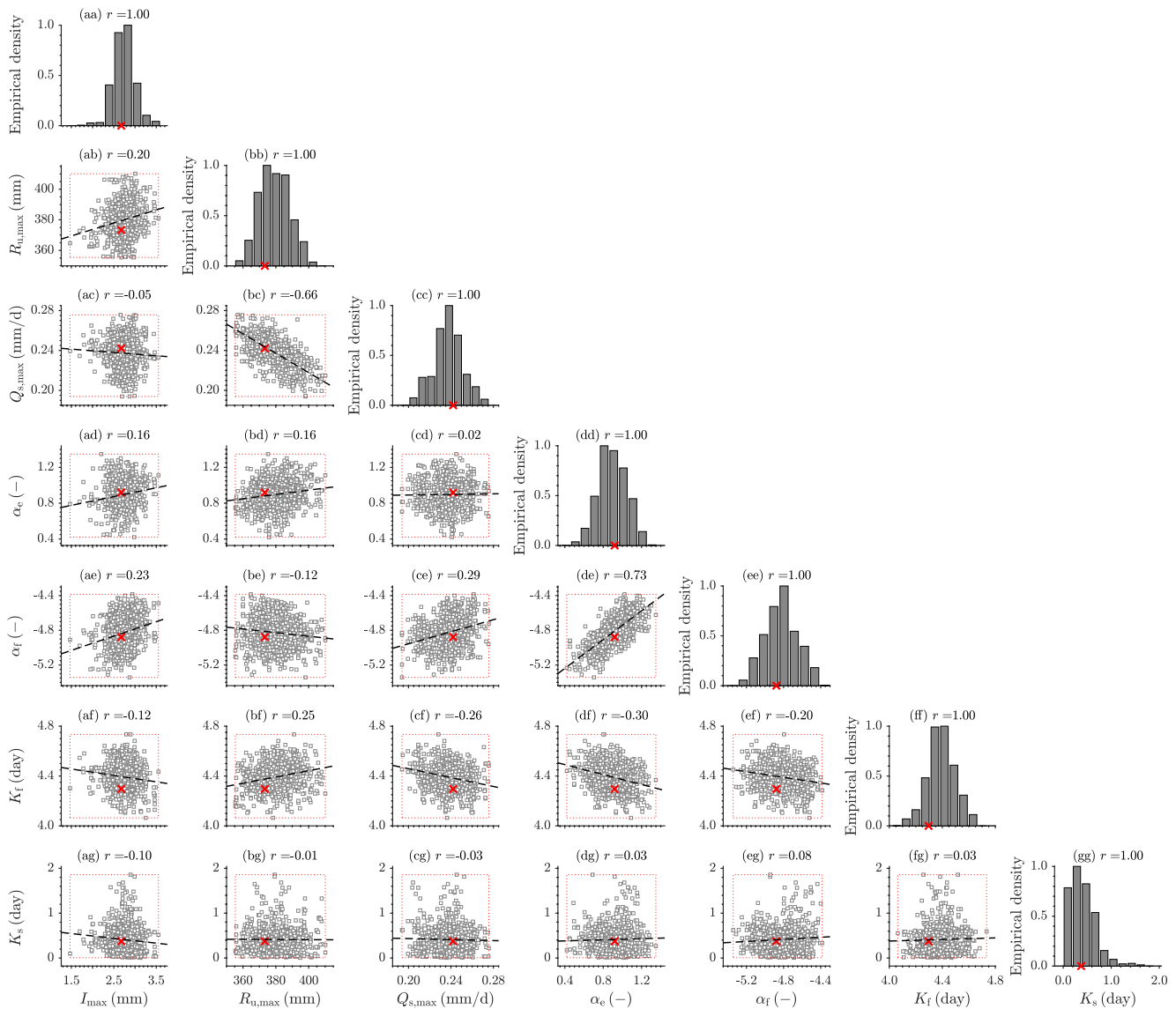


Figure 4. Matrix plot of the $DREAM_{(ZS)}$ -derived posterior realizations of the hmodel parameters. Histograms (main diagonal) of the marginal posterior hmodel parameter distributions and two-dimensional scatter plots (off-diagonal) of the bivariate posterior samples of hmodel parameter pairs. The solid dashed lines portray the least squares fit of a simple regression function to the sampled values. The red crosses correspond to the Maximum A Posteriori solution of the hmodel parameters. The red dotted lines depict the smallest rectangles that enclose the bivariate posterior samples of the $DREAM_{(ZS)}$ algorithm.

the marginal posterior parameter distributions are generally well described by a Gaussian distribution with small dispersion relative to the prior parameter ranges. The two-dimensional dot plots reveal the presence of correlation between the bivariate samples of the posterior distribution. The hmodel parameter pairs (α_e, α_f) and $(R_{u,max}, Q_{s,max})$ exhibit a particularly strong positive and negative correlation, respectively. This violates Sobol's method and necessitates the application of the variance/covariance-based model output decomposition of Equation 7.

The bivariate scatter plots confirm the presence of parameter correlation among the samples of the posterior distribution. What is less evident, however, is that the chain samples may also exhibit autocorrelation as the state of the Markov chain is duplicated each time a candidate point is rejected. Chain thinning hardly affects the results of our analysis and so did the use of only the unique posterior samples of the joint Markov chains. Thus, our findings are nearly invariant to the degree of autocorrelation and frequentness (=weight) of the posterior samples. Based on these grounds we use as input to probabilistic SA the $DREAM_{(ZS)}$ -sampled posterior realizations.

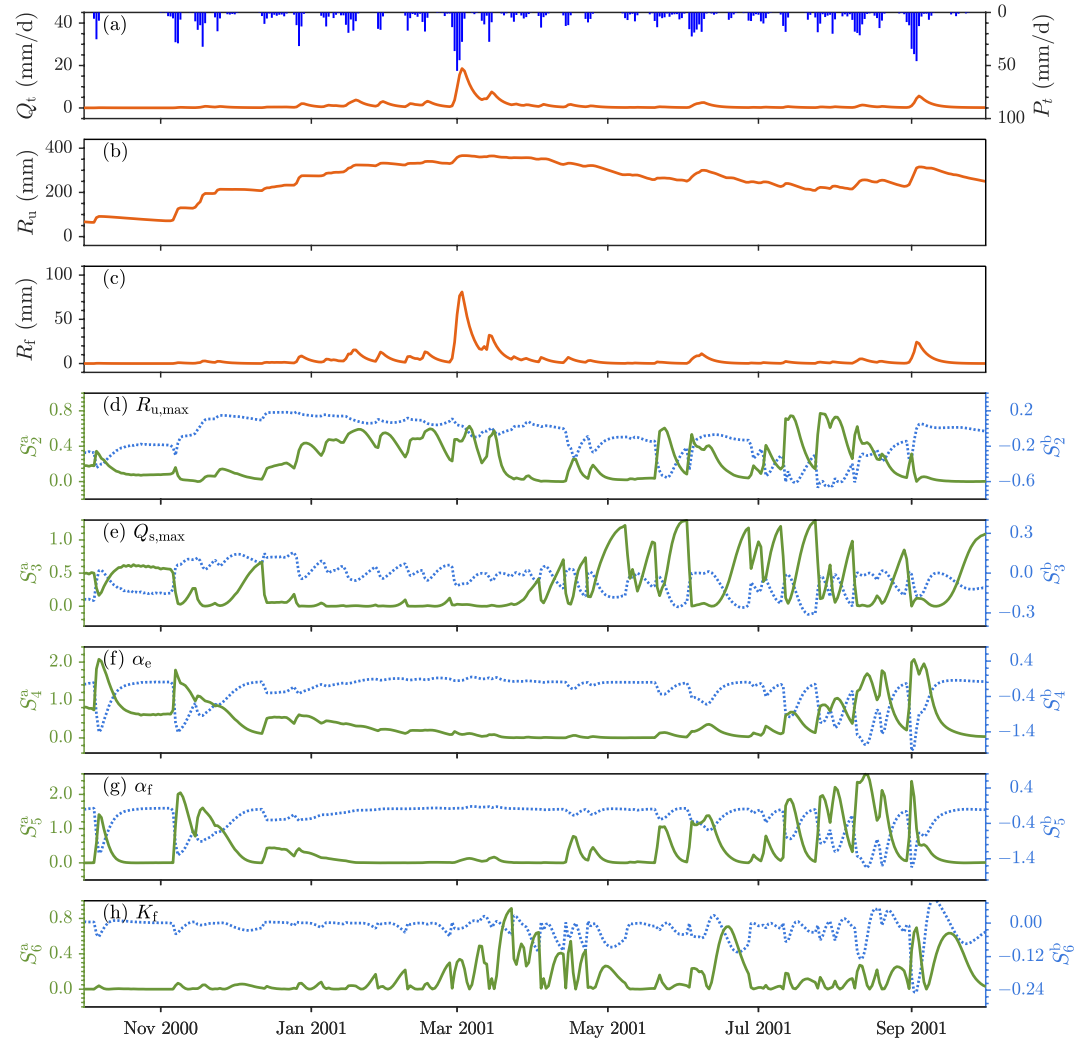


Figure 5. Precipitation, simulated discharge, hmodel state variables and $HDMR_{ext}$ estimates of structural and correlative sensitivity indices of hmodel parameters as function of time, where R_u and R_f are unsaturated zone and fast reservoir storage, respectively, the subscripts for the sensitivity indices, $\{2, 3, 4, 6, 7\}$, follow the same order of the parameters as they are presented in Table S3 in Supporting Information S1.

4.4.3. State/Time-Dependent Probabilistic Sensitivity Analysis

Figure 5 summarizes the results of $HDMR_{ext}$ and presents the evolution of the first-order structural, S_i^a , and correlative, S_i^b , sensitivity indices of (d) $R_{u,max}$, (e) $Q_{s,max}$, (f) α_e , (g) α_f , and (h) K_f during the period of 1 October 2000–30 September 2001. To understand the temporal dynamics of the indices, the parameter panels are preceded by time series plots of (a) rainfall and MAP simulated discharge, and the state of the (b) unsaturated and (c) fast reservoirs, respectively. The other hmodel parameters and state variables are displayed in Figure S8 in Supporting Information S1. We present only the first-order sensitivity indices as the higher-order terms exert only a minor control on the model output, y . This finding is supported by the values in Table S4 in Supporting Information S1 of the first- and second-order structural, correlative and total posterior sensitivity indices of the hmodel parameters on dry, moist and wet days.

The results in Figure 5 highlight several important findings. In the first place, notice that the structural (green) and correlative (blue) sensitivity indices exhibit considerable variations during the WY 2001. The temporal dynamics are particularly large for the structural sensitivity indices, S_i^a , with sudden and substantial fluctuations from one discharge value to the next. The traces of the correlative sensitivity indices appear smoother and more

compressed. Second, the magnitude of the structural sensitivity indices exceeds the correlative sensitivity indices. This is generally the case as the total main effect, $S_i^a + S_i^b$, is strictly positive, at least in theory. In fact, the correlative sensitivity indices are near zero for large portions of the 1-year discharge record. This is particularly true for (e) $Q_{s,max}$, (f) α_e , (g) α_f , and (h) K_f . Third, the structural and correlative sensitivity indices not only differ in magnitude and temporal dynamics but also in sign. The structural indices are strictly positive, whereas the correlative contribution usually takes on negative values. Thus, the correlative indices can substantially decrease the total sensitivity indices, $S_i = S_i^a + S_i^b$, of the input factors and, thus, main effects of the hmodel parameters. This is the reason so as to why the structural sensitivity indices of (e) $Q_{s,max}$, (f) α_e , and (g) α_f exceed unity for certain parts of the simulated hydrograph. This guarantees an exact variance decomposition with sum of the total sensitivity indices, $S_i = S_i^a + S_i^b$ that always amounts to unity for each successive discharge data point. Fourthly, the temporal behavior of the structural and/or correlative main effects does not seem to relate in an obvious manner to the hyetograph, hmodel simulated flow levels and/or wetness of the watershed as expressed in the state of the unsaturated and/or fast reservoir respectively. Altogether, the results demonstrate that parameter sensitivity is not an invariant property but rather varies dynamically depending on the model input, dominant processes, parameter values and state of the watershed. This is the underlying premise of AMALGAM (Vrugt & Robinson, 2007; Vrugt, Robinson, & Hyman, 2009).

The temporal behavior of the structural and correlative main effects of the hmodel parameters do not seem to correlate well with the hyetograph, simulated flow level and the watershed state. To investigate this in more detail, please consider Figure 6 which presents bivariate scatter plots of the first-order total sensitivity indices of the hmodel parameters and the cumulative rainfall amounts of the five preceding days, and simulated state of the interception, unsaturated zone, fast and slow reservoirs, respectively. The total main effects demonstrate a complex dependency on the cumulative rainfall and/or state of the watershed (Garambois et al., 2013; Herman et al., 2013; Huang et al., 2021; Pianosi & Wagener, 2016). This dependency is poorly characterized by a linear regression function and follows instead a more parabolic (dc, de, fc, fe) or hyperbolic (aa, ad, ca, cd, gd) relationship. The large data scatter confirms, however, that even such nonlinear relationships do not suffice to capture the temporal dynamics of the main effects. As a result, we resort to a more qualitative interpretation of the dotted plots. In doing so, we shall refer back to Figure S6 in Supporting Information S1 and the ensuing caption which summarizes the process equations of hmodel. The product-moment correlation coefficients and regression functions display a positive correlation between the states of the interception, unsaturated zone and fast reservoirs and the magnitude of the main effects of (ab) I_{max} , (bc) $R_{u,max}$, and (fd) K_f , respectively. This positive relationship is induced by the hmodel Equations which confirm that the impact of the parameters on the simulated discharge grows with the state of their respective reservoirs. The opposite is true for $Q_{s,max}$ and α_e , hence why we see a negative correlation in graph (cc) between the magnitude of the main effect of $Q_{s,max}$ and the state of the unsaturated zone reservoir. Furthermore, as the slow reservoir exerts only a minor control on the simulated discharge, the main effect of its recession constant, K_e , is close to zero during the 1-year discharge record in the bottom-right panel. Further research should explore the usefulness of the HDMR/HDMR_{ext} derived probabilistic sensitivity indices for improving hydrologic process representation.

4.4.4. Comparison of Prior and Posterior Sensitivity Indices With (In)dependent Samples

We investigate the impact of hmodel parameter correlation on the sensitivity indices derived from variance-based analysis of the simulated rainfall-discharge relationships. Figure 7 presents traces of the main-effects of each hmodel parameter derived from uniform (=independent) sampling of the prior, \mathbb{H}_0 , and posterior, \mathbb{H}_1 , hypercubes and the DREAM_(ZS) collection of dependent samples of the posterior distribution, $p(\mathbf{x}|\tilde{\mathbf{y}})$ (dash-dot green). In Figure S7 in Supporting Information S1 we display the corresponding ranges of the simulated discharge records. We use bullet points to discuss our main findings.

1. The first-order total sensitivity indices of the hmodel parameters derived from the dependent samples of the DREAM_(ZS) algorithm are generally in good agreement with the main effects of the independent samples of the posterior hypercube, \mathbb{H}_1 . For most hmodel parameters, I_{max} , $Q_{s,max}$, α_f , K_f , and K_s , the traces are markedly similar and typically only show differences in the magnitude of the sensitivity indices. This cannot be said for $R_{u,max}$ and α_e whose traces differ substantially, particularly during the second-half of the water year. This is a result of the relative strong correlation between ($R_{u,max}$, $Q_{s,max}$) and (α_e , α_f) so evidently visible in the scatter plots of Figure 4. Apparently, this correlation does not affect much the traces of $Q_{s,max}$ and α_f as it does $R_{u,max}$ and α_e .

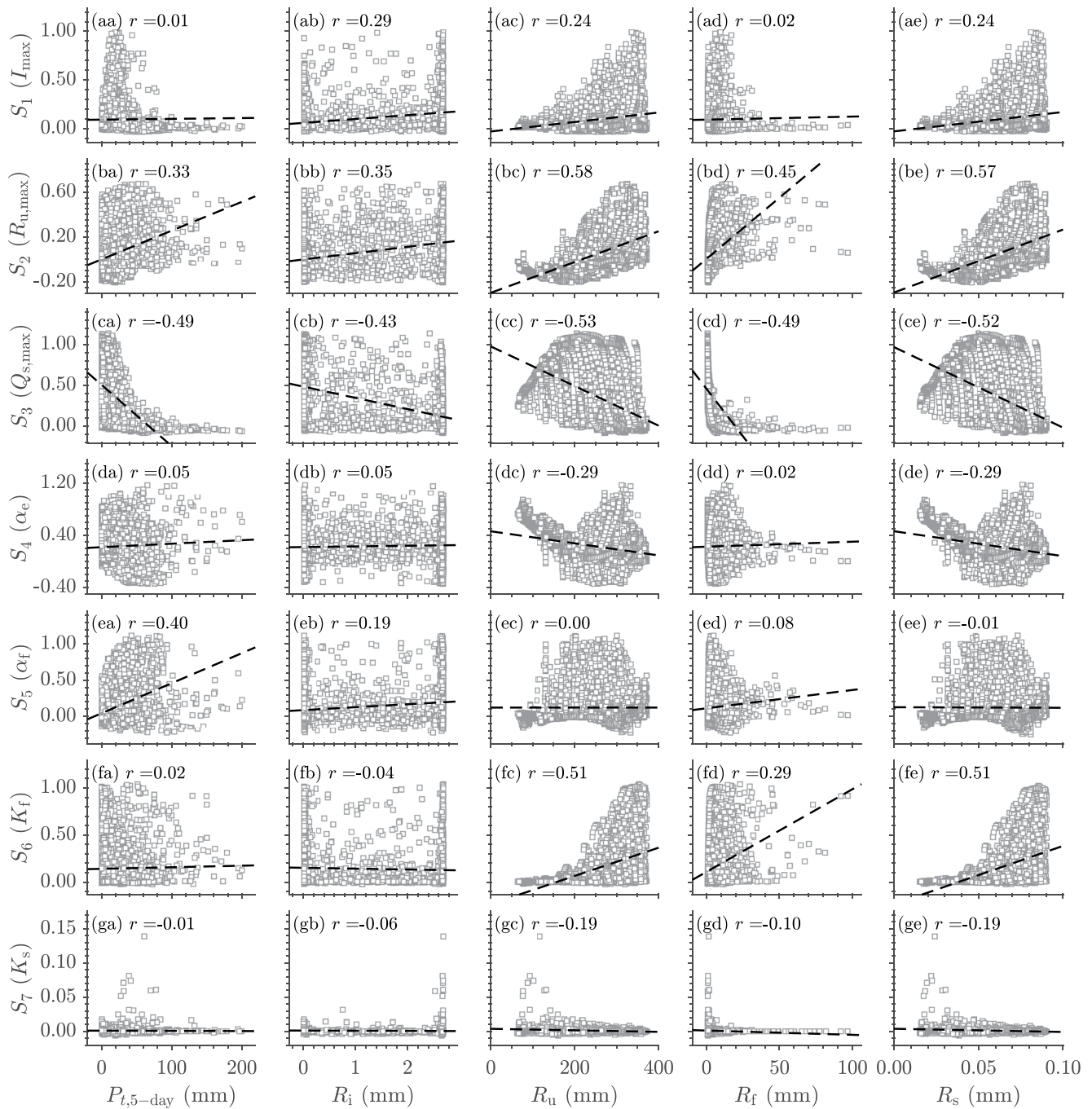


Figure 6. Dotty plots of the total main effect, $S_i = S_i^a + S_i^b$, of (a) I_{\max} , (b) $R_{u,\max}$, (c) $Q_{s,\max}$, (d) α_e , (e) α_f , (g) K_f (vertical axis) and (a) cumulative rainfall of the five preceding days and state of the (b) interception, (c) unsaturated zone, (d) fast, and (e) slow reservoirs (horizontal axis). In each graph we list Pearson's correlation coefficient of the data scatter and display the least squares fit of a regression function with slope and intercept.

2. A comparison of the main effects of the hmodel parameters in the posterior hypercube, \mathbb{H}_1 , and posterior distribution, $p(\mathbf{x}|\tilde{\mathbf{y}})$, reiterates that even mild correlation can change the order of input variable importance. This affects the classification of hmodel parameters into categories of having a very strong, mild, low and negligible impact on the simulated discharge.
3. The prior and posterior parameter sensitivities are very different from each other, not only in temporal variability but also in magnitude. Indeed, the prior main effects of (a) I_{\max} , (b) $R_{u,\max}$, (c) $Q_{s,\max}$, (d) α_e , (e) α_f ,

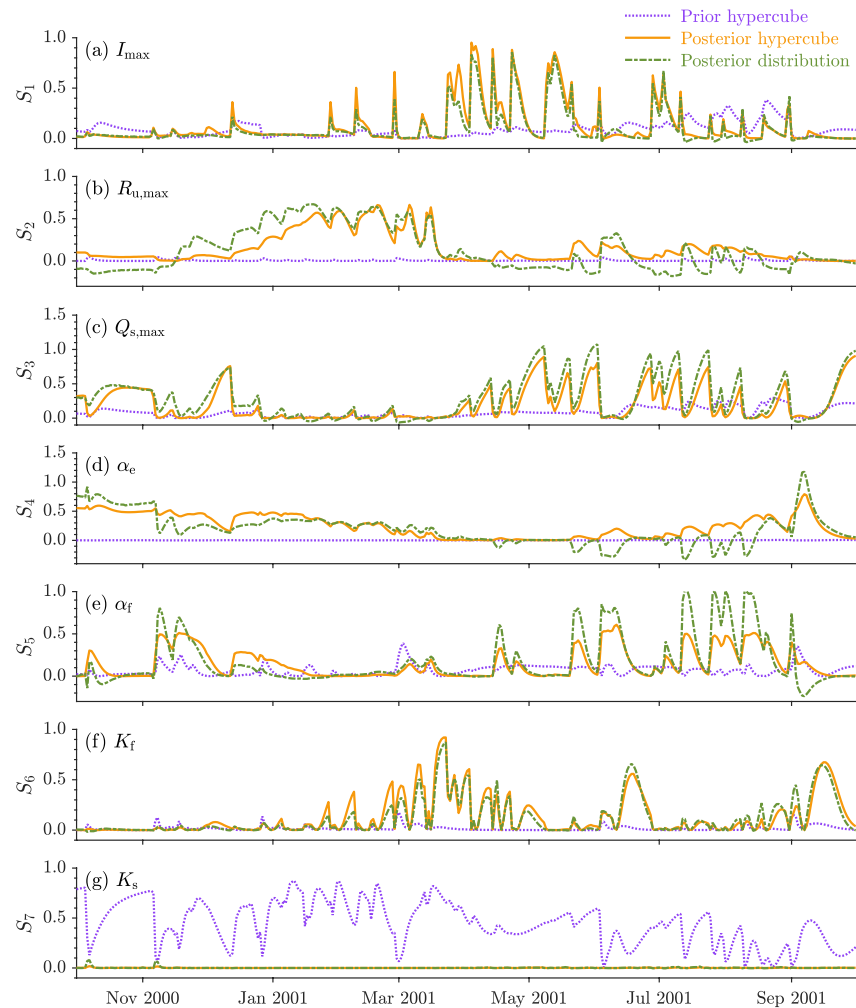


Figure 7. First-order total sensitivity indices, S_i , of (a) I_{\max} , (b) $R_{u,\max}$, (c) $Q_{s,\max}$, (d) α_e , (e) α_f , (g) K_f , and (h) K_s , for the 365-day discharge record using $N = 10,000$ independent samples of \mathbb{H}_0 (dotted purple) and \mathbb{H}_1 (solid orange) and ten-thousand posterior realizations of the DREAM_(ZS) algorithm (dash-dotted green).

and (f) K_f are much smaller than their posterior counterparts. In fact, their prior main effects are so small that one could classify these six parameters as insensitive. An opposite trend between the prior and posterior main-effect is found for the recession constant, K_s , of the slow reservoir. This parameter exerts a large and overarching control on the variance of the simulated discharges in the prior distribution but has a negligible impact on the discharge records of the posterior hypercube, \mathbb{H}_1 , with its main-effect, S_7 , reduced to almost zero. Thus, the Sobol' estimates of the main effects demonstrate a large change in parameter sensitivity when transitioning between the prior and posterior distribution. In this process, the contribution to the model output increases for all parameters but the recession constant K_s . These dynamic changes in parameter sensitivity will help guide search and optimization methods in pursuit of the optimum solution and/or posterior parameter distribution and determine the order of iterative refinement of the parameters. This impact of scale in SA is explored more explicitly by the VARS method of Razavi et al. (2019), Razavi and Gupta (2016) but obscures and/or clouds the ranking of input factor importance (Puy et al., 2021).

4. The total first-order sensitivity indices of $R_{u,\max}$, α_e , and α_f across the posterior distribution $p(\mathbf{x}|\bar{y})$ take on negative values on certain days of the water year. We have witnessed this before in the third case study and eluded to a simple analytic example in Text S1 in Supporting Information S1. Apparently, the correlative contributions of $R_{u,\max}$, α_e , and α_f to the simulated discharge are negative and larger in magnitude than their

respective structural contributions. Thus, on certain days these three parameters suppress the model output variance. We will revisit this finding in Section 5.

4.4.5. Surrogate Model Evaluation

To verify the accuracy of the HDMR_{ext} model output decomposition in Equation 25, for emulation of the rainfall-discharge relationship, we apply Algorithm 1 to the simulated discharge records of the $N = 10,000$ independent samples of the posterior hypercube, \mathbb{H}_1 . Figure 8a presents a frequency distribution of the sample variances, $s_e^2 = \frac{1}{N-l} \sum_{i=1}^N \epsilon_i^2$, of the N residuals of the daily surrogate models. A second-order decomposition was deemed sufficient to describe the data. As a result, $\delta = 2$, and the number of unknown expansion coefficients estimated with D-MORPH regression equals $l = 336$. The residual variances appear rather small. The median value of s_e^2 of about $2.1 \times 10^{-7} \text{ mm}^2/\text{d}^2$ is substantially smaller than the average within-day variance of $2.2 \times 10^{-3} \text{ mm}^2/\text{d}^2$ of the N discharges of the training samples. Thus, the variance decomposition should be close to exact. This inspires confidence in the sensitivity indices and surrogate model of HDMR_{ext} .

Next, we evaluate the accuracy of the surrogate models of Equation 25 for samples of \mathbf{x} that have not been used for D-MORPH coefficient estimation. Figure 8b displays a histogram of the variance of the surrogate model residuals computed for the ten-thousand posterior realizations of the $\text{DREAM}_{(\text{ZS})}$ algorithm. The use of dependent samples increases the median value of s_e^2 to about $1.1 \times 10^{-4} \text{ mm}^2/\text{d}^2$, about 20 times smaller than the average within-day variance of the discharges. Thus, within the hypercube, \mathbb{H}_1 , of the posterior parameter distribution, the second-order expansion of HDMR_{ext} accurately describes the rainfall-discharge relationship of the hmodel. Further research is warranted to determine the performance of the surrogate model over the domain, χ , of the prior hypercube, \mathbb{H}_0 .

5. Discussion and Conclusions

This paper has focused attention on probabilistic and/or variance-based SA with correlated input variables. We used the methodology of Li et al. (2010), called high dimensional model representation or HDMR, which is a generalization of ANOVA to dependent input factors and replaces the elementary functions of Sobol' by component functions with linear expansion coefficients. Under strict regularity conditions, the HDMR superposition of linear multiples of component functions will exactly decompose the variance of the model output and parse out the structural and correlative contributions of the input variables at first-, second-, third-, and higher-orders. We considered two different implementations of HDMR with variance/covariance decomposition of the model output up to the third-order. The default implementation of Li et al. (2010) uses cubic B-spline component functions and determines the expansion coefficients of each order sequentially through backfitting. A more powerful variant, HDMR_{ext} of Li and Rabitz (2012), implements orthonormalized polynomial component functions with extended bases and simultaneously infers the expansion coefficients of all orders using D-MORPH regression. This latter method guarantees hierarchical orthogonality of the component functions. The bootstrap method is implemented in both HDMR implementations to provide confidence intervals of the structural and correlative sensitivity indices at first, second- and third-order.

The power and usefulness of covariance-based SA was demonstrated by application of HDMR and HDMR_{ext} to a simple linear model and the well-known Ishigami function. These two studies confirmed that both methods can accurately decompose the variance of the model output and back out the structural and correlative contributions at first-, second-, and/or higher orders of the input factors. For independent input variables, the sensitivity indices of HDMR and HDMR_{ext} matched exactly their counterparts derived from Sobol'. For correlated input variables, the sum of the structural and correlative indices were in excellent agreement with the total sensitivity indices of modified Sobol' (Kucherenko et al., 2012).

The third case study considered application of HDMR to covariance-based SA of the water retention, $\theta(h)$, function of van Genuchten (1980). As the B-spline component functions in the default HDMR implementation did not guarantee an exact variance decomposition (as anticipated), we focused

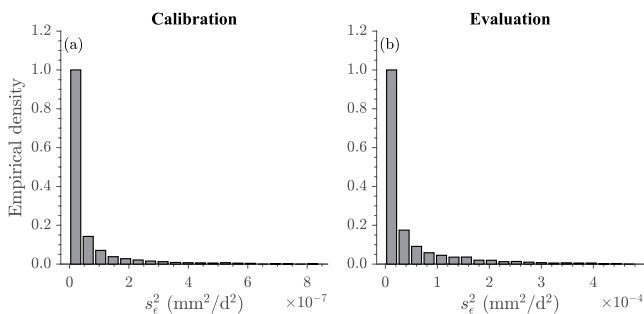


Figure 8. Histogram of the sample variances, s_e^2 , of the HDMR_{ext} surrogate model derived from (a) the training discharge records of $N = 10,000$ independent samples drawn from the posterior hypercube, \mathbb{H}_1 , and (b) evaluation discharge records of the collection of $N = 10,000$ posterior realizations of the $\text{DREAM}_{(\text{ZS})}$ algorithm.

our attention on HDMR_{ext} . The total effects of θ_s , θ_r , α_{vg} , and n_{vg} , appeared relatively unaffected by correlation and were in close agreement with the partial derivatives, $\partial\theta/\partial\theta_s$, $\partial\theta/\partial\theta_r$, $\partial\theta/\partial\alpha_{\text{vg}}$, and $\partial\theta/\partial n_{\text{vg}}$ of the van Genuchten (1980) soil water characteristic. Our results confirmed earlier findings of Vrugt et al. (2002) and Vrugt and ter Braak (2011) with respect to optimal experimental design of the hanging-column method.

Our fourth and last study considered probabilistic SA of the seven-parameter hmodel using the rainfall-discharge transformation of the Leaf River watershed in Mississippi. Posterior realizations sampled with the DREAM_(ZS) algorithm served as input to HDMR_{ext} to quantify sensitivity across the posterior distribution. The structural and correlative main effects of the hmodel parameters varied considerably during the water year in a manner that did not obviously relate to simulated flow level, catchment wetness, rainfall and potential evaporation. The total main effects of the posterior parameter distribution $p(\mathbf{x}|\bar{y})$ compared generally well with their Sobol' counterparts obtained from independent samples of the posterior hypercube, \mathbb{H}_1 . The main effects of two hmodel parameters changed considerably as a result of correlation. This affects the ranking of the main effects, and, thus, assessment of parameter importance.

The results of probabilistic SA of the hmodel appeared unaffected by the degree of autocorrelation among the posterior realizations and their frequency of appearance in the sampled Markov chains. Duplicate chain samples do not carry new information about the input-output relationship for the expansion coefficients, but do increase the weight that is attached to these samples in D-MORPH regression. Our results confirmed (not shown) that subsampling (thinning) of the Markov chains and/or the use of unique posterior samples only, did not change much the results of HDMR_{ext} . Thus, our covariance-based estimates of probabilistic sensitivity across the posterior hmodel parameter distribution appeared invariant to the exact distribution of the samples within this high-probability density region of the parameter space. If deemed necessary, one can fit a Gaussian mixture distribution to the posterior realizations using the methods presented in Volpi et al. (2017) and draw a large collection of samples from this distribution as input to probabilistic SA.

With only mild correlation among the hmodel parameters, the fourth and last case study did not do justice to the HDMR_{ext} methodology. When multivariate parameter dependencies are weak, the independence assumption of Sobol' will generally suffice to approximate the parameters' main effects across the posterior distribution. But then the posterior hypercube \mathbb{H}_1 must be delineated first from the samples of the posterior distribution. This two-step procedure will break down when multivariate parameter dependencies are strong and we must resort to covariance-based model decomposition with HDMR_{ext} to obtain robust variance-based estimates of the structural S_i^a correlative S_i^b and total S_i contributions of each input factor to the model output.

One can envision many other case studies for which an exact variance-covariance decomposition at first-, second- and higher-orders with HDMR_{ext} is a must for deriving sensical sensitivity indices. One can think of applications wherein input variables must satisfy a certain hierarchy, spatial arrangement or organization so as to not to violate system physics and/or modeling assumptions. For example, in surface hydrology one may wish to evaluate model sensitivity to spatiotemporal rainfall structure. In soil hydrology, one may want to determine the sensitivity of soil evaporation and/or root water uptake to surface temperature, humidity, wind-velocity and solar radiation. These meteorological variables are intimately related (Cascone et al., 2019) and treating them as independent would violate process physics and yield rather meaningless estimates of sensitivity. In all these cases, HDMR_{ext} can provide assistance in screening, model development and input prioritization. Furthermore, the HDMR_{ext} functional decomposition into linear multiples of component functions has application beyond (co)variance-based SA. This decomposition can serve as surrogate model in two-stage MCMC simulation to accelerate posterior exploration (Laloy et al., 2013; Paun & Husmeier, 2021). The total main effects of the parameters can also serve as guiding principle to continuously adapt search subspace in DREAM Suite (Vrugt, 2016) in pursuit of the posterior parameter distribution.

While HDMR_{ext} is a powerful addition to the hydrologists' arsenal of SA methods, this method can pose considerable computational challenges. The number of expansion coefficients l and, thus, number of columns of the design matrix Φ in Equation 32 increases polynomially with the number of input variables d , polynomial degree p of the component functions and order δ of the functional decomposition in Equation 25. This complicates the application of HDMR_{ext} to parameter-rich models. For example, if $d = 10$, $p = 3$, and $\delta = 2$, we have $l = 705$ unknown expansion coefficients. For a third-order decomposition $\delta = 3$ this value increases to $l = 8,265$. The rapid growth of the number of expansion coefficients with d , p , and/or δ can pose problems with the storage of the design matrix Φ in random-access memory (RAM). Suppose, for example, we use a ratio $N/l = 5$, then a stand-alone PC with 16 gigabytes of RAM can handle a maximum of $l = 20,000$ expansion coefficients before depleting the memory. If we

assume default settings of $p = 3$, then a second-order model decomposition would reach this number of expansion coefficients for $d = 51$. For $\delta = 3$ the maximum number of input variables is limited only to $d = 13$. Thus, there are significant advantages to keeping the order δ of the HDMR model decomposition and polynomial degree p of the component functions as low as possible. Fortunately, most physics-based models are of type A or B according to the functional classification of Kucherenko et al., 2011 and, thus, support a second-order model decomposition (Falchi et al., 2018; Rabitz & Aliş, 1999; Ratto et al., 2007; Shereena & Rao, 2019). If the input-output relationship warrants use of a higher-order decomposition (e.g., Huang et al., 2021) then distributed-memory and/or multi-core computer system is required to implement ANCOVA-based model decomposition with HDMR_{ext} .

As closing remark we comment on the negative values of the total- and main effect sensitivity indices, S_i^T and S_p , respectively, in case studies three and four. This non-intuitive finding is not a numerical artifact but the result of applying the total law of covariance as is demonstrated in Text S1 and Table S5 in Supporting Information S1 with a simple analytic example. This study confirms that a parameter's negative correlative contribution can exceed in magnitude its structural contribution. Thus, some input variables can suppress the model output variance by promoting, for example, negative feed backs in the model and driving the output to an equilibrium.

Data Availability Statement

Our general-purpose MATLAB toolboxes of HDMR and HDMR_{ext} (V. 1.0, executable files), “MATLAB_code_HDMR_V1.0” and “MATLAB_code_HDMR_EXT_V1.0,” along with the case studies of this paper can be downloaded from Zenodo at <https://doi.org/10.5281/zenodo.7478901> (Vrugt et al., 2022) with Creative Commons Attribution 4.0 International license. The toolboxes are also available from GitHub at <https://github.com/jaspervrugt/HDMR> licensed under MIT. The CAMELS data set used in the last study is described in Newman et al. (2015) and available at <https://dx.doi.org/10.5065/D6MW2F4D> (Newman et al., 2014) with Creative Commons Attribution 4.0 International license.

Acknowledgments

The constructive comments of the reviewers have substantially improved our manuscript. We would like to thank Dr. Genyuan Li for diligently answering our technical questions about the HDMR methodology. We also thank Dr. Debora Yumi de Oliveira for providing the posterior samples and simulations of the hmodel parameters.

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>
- Bennett, F. R., & Fentie, B. (2017). Sensitivity analysis of constituent generation parameters of an integrated hydrological and water quality model using a GMDH polynomial neural network. In *MODSIM2017, 22nd international congress on modelling and simulation, modelling and simulation society of Australia and New Zealand, December, 1899–1905*.
- Beven, K., & Binley, A. (1992). The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes*, 6(3), 279–298. <https://doi.org/10.1002/hyp.3360060305>
- Borgonovo, E., Lu, X., Plishcke, E., Rakovec, O., & Hill, M. C. (2017). Making the most out of a hydrological model data set: Sensitivity analyses to open the model black-box. *Water Resources Research*, 53(9), 7933–7950. <https://doi.org/10.1002/2017WR020767>
- Brooks, S., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational & Graphical Statistics*, 7(4), 434–455. <https://doi.org/10.2307/1390675>
- Brunner, M. I., Melsen, L. A., Newman, A. J., Wood, A. W., & Clark, M. P. (2020). Future streamflow regime changes in the United States: Assessment using functional classification. *Hydrology and Earth System Sciences*, 24(8), 3951–3966. <https://doi.org/10.5194/hess-24-3951-2020>
- Carsel, R. F., & Parrish, R. S. (1988). Developing joint probability distributions of soil water retention characteristics. *Water Resources Research*, 24(5), 755–769. <https://doi.org/10.1029/WR024i005p00755>
- Cascone, S., Coma, J., Gagliano, A., & Perez, G. (2019). The evapotranspiration process in green roofs: A review. *Building and Environment*, 147, 337–355. <https://doi.org/10.1016/j.buildenv.2018.10.024>
- Chastaing, G., Gamboa, F., & Prieur, C. (2015). Generalized Sobol' sensitivity indices for dependent variables: Numerical methods. *Journal of Statistical Computation and Simulation*, 85(7), 1306–1333. <https://doi.org/10.1080/00949655.2014.960415>
- Cox, M. G. (1972). The numerical evaluation of B-splines. *IMA Journal of Applied Mathematics*, 10(2), 134–149. <https://doi.org/10.1093/imamat/10.2.134>
- Cukier, R. I., Fortuin, C. M., Shuler, K. E., Petschek, A. G., & Schaibly, J. H. (1973). Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I theory. *The Journal of Chemical Physics*, 59(8), 3873–3878. <https://doi.org/10.1063/1.1680571>
- De Boor, C. (1972). On calculating with B-splines. *Journal of Approximation Theory*, 6(1), 50–62. [https://doi.org/10.1016/0021-9045\(72\)90080-9](https://doi.org/10.1016/0021-9045(72)90080-9)
- Do, N. C., & Razavi, S. (2020). Correlation effects? A major but often neglected component in sensitivity and uncertainty analysis. *Water Resources Research*, 56(3), e2019WR025436. <https://doi.org/10.1029/2019WR025436>
- Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC Press.
- Falchi, A., Minisci, E., Kubicek, M., Vasile, M., & Lemmens, S. (2018). HDMR-based sensitivity analysis and uncertainty quantification of GOCE aerodynamics using DSMC. *Stardust Final Conference*, 301–323. <https://doi.org/10.1007/978-3-319-69956-1>
- Garambois, P. A., Roux, H., Larnier, K., Castaings, W., & Dartus, D. (2013). Characterization of process-oriented hydrologic model behavior with temporal sensitivity analysis for flash floods in Mediterranean catchments. *Hydrology and Earth System Sciences*, 17(6), 2305–2322. <https://doi.org/10.5194/hess-17-2305-2013>
- Gelman, A., & Rubin, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4), 457–511. <https://doi.org/10.1214/ss/1177011136>
- Golub, G. H., & Van Loan, C. F. (1996). *Matrix computations* (3rd ed.). Johns Hopkins.

- Gomes, G. J. C., Vrugt, J. A., Vargas, E. A., Camargo, J. T., Velloso, R. Q., & van Genuchten, M. T. (2017). The role of uncertainty in bedrock depth and hydraulic properties on the stability of a variably-saturated slope. *Computers and Geotechnics*, 88, 222–241. <https://doi.org/10.1016/j.compgeo.2017.03.016>
- Herman, J. D., Reed, P. M., & Wagener, T. (2013). Time-varying sensitivity analysis clarifies the effects of watershed model formulation on model behavior. *Water Resources Research*, 49(3), 1400–1414. <https://doi.org/10.1002/wrcr.20124>
- Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, 19(3), 293–325. <https://doi.org/10.1007/978-1-4612-0919-5>
- 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)
- Hooker, G. (2007). Generalized functional ANOVA diagnostics for high-dimensional functions of dependent variables. *Journal of Computational & Graphical Statistics*, 16(3), 709–732. <https://doi.org/10.1198/106186007X237892>
- Huang, C., Tong, J., & Ye, M. (2021). Global sensitivity analysis for a prediction model of soil solute transfer into surface runoff. *Journal of Hydrology*, 599, 126342. <https://doi.org/10.1016/j.jhydrol.2021.126342>
- Ishigami, T., & Homma, T. (1990). An importance quantification technique in uncertainty analysis for computer models. In *First international symposium on uncertainty modeling and analysis* (pp. 398–403). IEEE. <https://doi.org/10.1109/ISUMA.1990.151285>
- Jacques, J., Lavergne, C., & Devictor, N. (2006). Sensitivity analysis in presence of model uncertainty and correlated inputs. *Reliability Engineering & System Safety*, 91(10), 1126–1134. <https://doi.org/10.1016/j.res.2005.11.047>
- Khare, Y. P., Martinez, C. J., & Muoz-Carpena, R. (2013). Parameter variability and drought models: A study using the agricultural reference index for drought (ARID). *Agronomy Journal*, 105(5), 1417–1432. <https://doi.org/10.2134/agronj2013.0167>
- Khorashadi Zadeh, F., Nossent, J., Sarrazin, F., Pianosi, F., van Griensven, A., Wagener, T., & Bauwens, W. (2017). Comparison of variance-based and moment-independent global sensitivity analysis approaches by application to the SWAT model. *Environmental Modelling & Software*, 91, 210–222. <https://doi.org/10.1016/j.envsoft.2017.02.001>
- Kiparissides, A., Kucherenko, S. S., Mantalaris, A., & Pistikopoulos, E. N. (2009). Global sensitivity analysis challenges in biological systems modeling. *Industrial & Engineering Chemistry Research*, 48(15), 7168–7180. <https://doi.org/10.1021/ie900139x>
- Kucherenko, S., Albrecht, D., & Saltelli, A. (2015). Exploring multi-dimensional spaces: A comparison of Latin hypercube and quasi Monte Carlo sampling techniques. *arXiv preprint*. arXiv:1505.02350. <https://doi.org/10.48550/arXiv.1505.02350>
- Kucherenko, S., Feil, B., Shah, N., & Mauntz, W. (2011). The identification of model effective dimensions using global sensitivity analysis. *Reliability Engineering & System Safety*, 96(4), 440–449. <https://doi.org/10.1016/j.res.2010.11.003>
- Kucherenko, S., Klymenko, O. V., & Shah, N. (2017). Sobol' indices for problems defined in nonrectangular domains. *Reliability Engineering & System Safety*, 167, 218–231. <https://doi.org/10.1016/j.res.2017.06.001>
- Kucherenko, S., & Song, S. (2016). Derivative-based global sensitivity measures and their link with Sobol' sensitivity indices. In *Monte Carlo and quasi-Monte Carlo methods* (pp. 455–469). Springer. <https://doi.org/10.1007/978-3-319-33507-0>
- 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>
- Laloy, E., Rogiers, B., Vrugt, J. A., Mallants, D., & Jacques, D. (2013). Efficient posterior exploration of a high-dimensional groundwater model from two-stage Markov chain Monte Carlo simulation and polynomial chaos expansion. *Water Resources Research*, 49(5), 2664–2682. <https://doi.org/10.1002/wrcr.20226>
- Li, G., & Rabitz, H. (2010). D-MORPH regression: Application to modeling with unknown parameters more than observation data. *Journal of Mathematical Chemistry*, 48(4), 1010–1035. <https://doi.org/10.1007/s10910-010-9722-2>
- Li, G., & Rabitz, H. (2012). General formulation of HDMR component functions with independent and correlated variables. *Journal of Mathematical Chemistry*, 50(1), 99–130. <https://doi.org/10.1007/s10910-011-9898-0>
- 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>
- Li, G., Rosenthal, C., & Rabitz, H. (2001). High dimensional model representations. *The Journal of Physical Chemistry A*, 105(33), 7765–7777. <https://doi.org/10.1021/jp010450t>
- Liu, Y., & Gupta, H. V. (2007). Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. *Water Resources Research*, 43(W07401), 1–18. <https://doi.org/10.1029/2006WR005756>
- 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., Taranto, 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>
- 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.res.2011.08.008>
- Marrel, A., Iooss, B., Laurent, B., & Roustant, O. (2009). Calculations of Sobol' indices for the Gaussian process metamodel. *Reliability Engineering & System Safety*, 94(3), 742–751. <https://doi.org/10.1016/j.res.2008.07.008>
- McKay, M. D. (1995). *Evaluating prediction uncertainty (technical report)*. Los Alamos National Laboratory.
- McKay, M. D. (1997). Nonparametric variance-based methods of assessing uncertainty importance. *Reliability Engineering & System Safety*, 57(3), 267–279. [https://doi.org/10.1016/S0951-8320\(97\)00039-2](https://doi.org/10.1016/S0951-8320(97)00039-2)
- Mockler, E. M., O'Loughlin, F. E., & Bruen, M. (2016). Understanding hydrological flow paths in conceptual catchment models using uncertainty and sensitivity analysis. *Computers & Geosciences*, 90, 66–77. <https://doi.org/10.1016/j.cageo.2015.08.015>
- Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., et al. (2015). Development of a large-sample watershed-scale hydro-meteorological data set for the contiguous USA: Data set characteristics and assessment of regional variability in hydrologic model performance. *Hydrology and Earth System Sciences*, 19(1), 209–223. <https://doi.org/10.5194/hess-19-209-2015>
- Newman, A. J., Sampson, K., Clark, M. P., Bock, A., Viger, R. J., & Blodgett, D. (2014). A large sample watershed-scale hydrometeorological dataset for the contiguous USA [Dataset]. Boulder, CO: UCAR/NCAR. <https://doi.org/10.5065/D6MW2F4D>
- Oakley, J. E., & O'Hagan, A. (2004). Probabilistic sensitivity analysis of complex models: A Bayesian approach. *Journal of the Royal Statistical Society: Series B*, 66(3), 751–769. <https://doi.org/10.1111/j.1467-9868.2004.05304.x>
- Owen, A. B. (2005). Multidimensional variation for quasi-Monte Carlo. In J. Fan & G. Li (Eds.), *International conference on statistics in honour of professor kai-tai fang's 65th birthday* (pp. 49–74).
- Pan, F., Ye, M., Zhu, J., Wu, Y., Hu, B. X., & Yu, Z. (2009). Numerical evaluation of uncertainty in water retention parameters and effect on predictive uncertainty. *Vadose Zone Journal*, 8(1), 158–166. <https://doi.org/10.2136/vzj2008.0092>

- Pan, F., Zhu, J., Ye, M., Pachepsky, Y. A., & Wu, Y. (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>
- Paun, L. M., & Husmeier, D. (2021). Markov chain Monte Carlo with Gaussian processes for fast parameter estimation and uncertainty quantification in a 1D fluid-dynamics model of the pulmonary circulation. *International Journal for Numerical Methods in Biomedical Engineering*, 37(2), e3421. <https://doi.org/10.1002/cnm.3421>
- Penrose, R. (1955). A generalized inverse for matrices. *Proceedings of the Cambridge Philosophical Society*, 51(3), 406–413. <https://doi.org/10.1017/S0305004100030401>
- Pianosi, F., & Wagener, T. (2016). Understanding the time-varying importance of different uncertainty sources in hydrological modelling using global sensitivity analysis. *Hydrological Processes*, 30(22), 3991–4003. <https://doi.org/10.1002/hyp.10968>
- Puy, A., Piano, S. L., & Saltelli, A. (2021). Is VARS more intuitive and efficient than Sobol' indices. *Environmental Modelling & Software*, 137, 104960. <https://doi.org/10.1016/j.envsoft.2021.104960>
- Rabitz, H., & Aliş, Ö. F. (1999). General foundations of high-dimensional model representations. *Journal of Mathematical Chemistry*, 25(2), 197–233. <https://doi.org/10.1023/A:1019188517934>
- Ratto, M., Pagano, A., & Young, P. (2007). State dependent parameter metamodelling and sensitivity analysis. *Computer Physics Communications*, 177(11), 863–876. <https://doi.org/10.1016/j.cpc.2007.07.011>
- Razavi, S., & Gupta, H. V. (2016). A new framework for comprehensive, robust, and efficient global sensitivity analysis: 1. Theory. *Water Resources Research*, 52(1), 423–439. <https://doi.org/10.1002/2015WR017558>
- 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>
- Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., & Tarantola, S. (2010). Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2), 259–270. <https://doi.org/10.1016/j.cpc.2009.09.018>
- Saltelli, A., Bamber, G., Bruno, I., Charters, E., Di Fiore, M., Didier, E., et al. (2020). Five ways to ensure that models serve society: A manifesto. *Nature*, 582(7813), 482–484. <https://doi.org/10.1038/d41586-020-01812-9>
- Saltelli, A., & Tarantola, S. (2002). On the relative importance of input factors in mathematical models: Safety assessment for nuclear waste disposal. *Journal of the American Statistical Association*, 97(459), 702–709. <https://doi.org/10.1198/016214502388618447>
- Schoups, G., & Vrugt, J. A. (2010). A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. *Water Resources Research*, 46(10), W10531. <https://doi.org/10.1029/2009WR008933>
- Schoups, G., Vrugt, J. A., Fenicia, F., & van de Giesen, N. C. (2010). Corruption of accuracy and efficiency of Markov chain Monte Carlo simulation by inaccurate numerical implementation of conceptual hydrologic models. *Water Resources Research*, 46(10), W10530. <https://doi.org/10.1029/2009WR008648>
- Shereena, O. A., & Rao, B. N. (2019). HDMR-based Bayesian structural system identification. In *Recent advances in structural engineering* (Vol. 1, pp. 453–464). Springer. <https://doi.org/10.1007/978-981-13-0362-3>
- Sobol', I. M. (1967). On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics*, 7(4), 86–112. [https://doi.org/10.1016/0041-5553\(67\)90144-9](https://doi.org/10.1016/0041-5553(67)90144-9)
- Sobol', I. M. (1990). On sensitivity estimation for nonlinear mathematical models. *Matematicheskoe Modelirovanie*, 2(1), 112–118.
- Sobol', I. M. (1993). Sensitivity estimates for nonlinear mathematical models. *Mathematical Modelling and Computational Experiment*, 1(4), 407–414.
- Sobol', I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(1), 271–280. [https://doi.org/10.1016/S0378-4754\(00\)00270-6](https://doi.org/10.1016/S0378-4754(00)00270-6)
- 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>
- Storlie, C. B., Swiler, L. P., Helton, J. C., & Sallaberry, C. J. (2009). Implementation and evaluation of nonparametric regression procedures for sensitivity analysis of computationally demanding models. *Reliability Engineering & System Safety*, 94(11), 1735–1763. <https://doi.org/10.1016/j.res.2009.05.007>
- Tarantola, S., Gatelli, D., & Mara, T. (2006). Random balance designs for the estimation of first order global sensitivity indices. *Reliability Engineering & System Safety*, 91(6), 717–727. <https://doi.org/10.1016/j.res.2005.06.003>
- 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>
- ter Braak, C. J. F. (2006). A Markov chain Monte Carlo version of the genetic algorithm differential evolution: Easy Bayesian computing for real parameter spaces. *Statistics and Computing*, 16(3), 239–249. <https://doi.org/10.1007/s11222-006-8769-1>
- van Genuchten, M. T. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society of America Journal*, 44(5), 892–898. <https://doi.org/10.2136/sssaj1980.03615995004400050002x>
- Volpi, E., Schoups, G., Firmani, G., & Vrugt, J. A. (2017). Sworn testimony of the model evidence: Gaussian mixture importance (GAME) sampling. *Water Resources Research*, 53(7), 6133–6158. <https://doi.org/10.1002/2016WR020167>
- 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>
- Vrugt, J. A., Bouten, W., Gupta, H. V., & Sorooshian, S. (2002). Toward improved identifiability of hydrologic model parameters: The information content of experimental data. *Water Resources Research*, 38(12), 48-1–48-13. <https://doi.org/10.1029/2001WR001118>
- Vrugt, J. A., Diks, C. G. H., Gupta, H. V., Bouten, W., & Verstraten, J. M. (2005). Improved treatment of uncertainty in hydrologic modeling: Combining the strengths of global optimization and data assimilation. *Water Resources Research*, 41(1), W01017. <https://doi.org/10.1029/2004WR003059>
- Vrugt, J. A., Gao, Y., & Sahin, A. (2022). MATLAB toolboxes of high dimensional model representation (HDMR) with uncorrelated and/or correlated input variables (version 1.0) [Software]. Zenodo. <https://doi.org/10.5281/zenodo.7478901>
- Vrugt, J. A., & Robinson, B. A. (2007). Improved evolutionary optimization from genetically adaptive multimethod search. *Proceedings of the National Academy of Sciences*, 104(3), 708–711. <https://doi.org/10.1073/pnas.0610471104>
- Vrugt, J. A., Robinson, B. A., & Hyman, J. M. (2009). Self-Adaptive multimethod search for global optimization in real-parameter spaces. *IEEE Transactions on Evolutionary Computation*, 13(2), 243–259. <https://doi.org/10.1109/TEVC.2008.924428>
- Vrugt, J. A., ter Braak, C., Diks, C., Robinson, B. A., Hyman, J. M., & Higdon, D. (2009). Accelerating Markov chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Sciences and Numerical Simulation*, 10(3), 273–290. <https://doi.org/10.1515/IJNSNS.2009.10.3.273>

- Vrugt, J. A., ter Braak, C. J. F., Clark, M. P., Hyman, J. M., & Robinson, B. A. (2008). Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation. *Water Resources Research*, *44*(12), W00B09. <https://doi.org/10.1029/2007WR006720>
- Vrugt, J. A., & ter Braak, C. J. K. (2011). DREAM(D): An adaptive Markov chain Monte Carlo simulation algorithm to solve discrete, noncontinuous, and combinatorial posterior parameter estimation problem. *Hydrology and Earth System Sciences*, *5*(12), 3701–3713. <https://doi.org/10.5194/hess-15-3701-2011>
- Wang, A., Pianosi, F., & Wagener, T. (2020). Technical report—Methods: A diagnostic approach to analyze the direction of change in model outputs based on global variations in the model inputs. *Water Resources Research*, *56*(8), e2020WR027153. <https://doi.org/10.1029/2020WR027153>
- Wang, A., & Solomatine, D. P. (2019). Practical experience of sensitivity analysis: Comparing six methods, on three hydrological models, with three performance criteria. *Water*, *11*(5), 1062. <https://doi.org/10.3390/w11051062>
- Wang, H., Chen, L., Ye, F., & Chen, L. (2017). Global sensitivity analysis for fiber reinforced composite fiber path based on D-MORPH-HDMR algorithm. *Structural and Multidisciplinary Optimization*, *56*(3), 697–712. <https://doi.org/10.1007/s00158-017-1681-9>
- Wang, H., Gong, W., Duan, Q., & Di, Z. (2020). Evaluation of parameter interaction effect of hydrological models using the sparse polynomial chaos (SPC) method. *Environmental Modelling & Software*, *125*, 104612. <https://doi.org/10.1016/j.envsoft.2019.104612>
- Xu, C. (2013). Decoupling correlated and uncorrelated uncertainty contributions for nonlinear models. *Applied Mathematical Modelling*, *37*(24), 9950–9969. <https://doi.org/10.1016/j.apm.2013.05.036>
- Xu, C., & Gertner, G. Z. (2008a). A general first-order global sensitivity analysis method. *Reliability Engineering & System Safety*, *93*(7), 1060–1071. <https://doi.org/10.1016/j.ress.2007.04.001>
- Xu, C., & Gertner, G. Z. (2008b). 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>
- Yuan, Y., Khare, Y., Wang, X., Parajuli, P. B., Kisekka, I., & Finsterle, S. (2015). Hydrologic and water quality models: Sensitivity. *Transactions of the ASABE*, *58*(6), 1721–1744. <https://doi.org/10.13031/trans.58.10611>
- Zeilew, M. B., & Alfredeisen, K. (2013). Sensitivity-guided evaluation of the HBV hydrological model parameterization. *Journal of Hydroinformatics*, *15*(3), 967–990. <https://doi.org/10.2166/hydro.2012.011>
- 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>
- Ziehn, T., & Tomlin, A. S. (2008). A global sensitivity study of sulfur chemistry in a premixed methane flame model using HDMR. *International Journal of Chemical Kinetics*, *40*(11), 742–753. <https://doi.org/10.1002/kin.20367>
- Zuniga, M. M., Kucherenko, S., & Shah, N. (2013). Metamodeling with independent and dependent inputs. *Computer Physics Communications*, *184*(6), 1570–1580. <https://doi.org/10.1016/j.cpc.2013.02.005>

References From the Supporting Information

- Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., & Loumagne, C. (2005). Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2—Towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling. *Journal of Hydrology*, *303*(1–4), 290–306. <https://doi.org/10.1016/j.jhydrol.2004.08.026>
- Schaap, M. G., & Bouten, W. (1996). Modeling water retention curves of sandy soils using neural networks. *Water Resources Research*, *32*(10), 3033–3040. <https://doi.org/10.1029/96WR02278>
- Schaap, M. G., Leij, F. J., & van Genuchten, M. T. (2001). ROSETTA: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *Journal of Hydrology*, *251*(3), 163–176. [https://doi.org/10.1016/S0022-1694\(01\)00466-8](https://doi.org/10.1016/S0022-1694(01)00466-8)