Environmental Modelling and Software 112 (2019) 95–107

Contents lists available at ScienceDirect

Environmental Modelling & Software

journal homepage: www.elsevier.com/locate/envsoft

VARS-TOOL: A toolbox for comprehensive, efficient, and robust sensitivity and uncertainty analysis

Saman Razavi[a,b,c,*], Razi Sheikholeslami[a,b], Hoshin V. Gupta[d], Amin Haghnegahdar[a,b]

[a] Global Institute for Water Security, University of Saskatchewan, Saskatoon, Saskatchewan, Canada
[b] School of Environment and Sustainability, University of Saskatchewan, Saskatoon, Saskatchewan, Canada
[c] Department of Civil, Geological, and Environmental Engineering, University of Saskatchewan, Saskatoon, Saskatchewan, Canada
[d] Department of Hydrology & Atmospheric Sciences, The University of Arizona, Tucson, AZ, USA

ARTICLE INFO

Keywords:
Global sensitivity analysis
Uncertainty analysis
Variogram analysis of response surface (VARS)
Sobol'
Morris
Progressive Latin hypercube sampling (PLHS)
Dynamical systems models
Sensitivity indices
Performance metrics

ABSTRACT

VARS-TOOL is a software toolbox for sensitivity and uncertainty analysis. Developed primarily around the "*Variogram Analysis of Response Surfaces*" framework, VARS-TOOL adopts a multi-method approach that enables simultaneous generation of a range of sensitivity indices, including ones based on derivative, variance, and variogram concepts, from a *single* sample. Other special features of VARS-TOOL include (1) novel tools for time-varying and time-aggregate sensitivity analysis of dynamical systems models, (2) highly efficient sampling techniques, such as Progressive Latin Hypercube Sampling (PLHS), that maximize robustness and rapid convergence to stable sensitivity estimates, (3) factor grouping for dealing with high-dimensional problems, (4) visualization for monitoring stability and convergence, (5) model emulation for handling model crashes, and (6) an interface that allows working with any model in any programming language and operating system. As a test bed for training and research, VARS-TOOL provides a set of mathematical test functions and the (dynamical) HBV-SASK hydrologic model.

Software availability

The VARS-TOOL software package as well as the HBV-SASK hydrologic model and its case studies can be downloaded from www.vars-tool.com and be used free of charge for non-commercial purposes.

1. Introduction

Earth and environmental systems models are widely employed for the simulation of complex physical processes that comprise the Earth's natural and engineered systems (Bennett et al., 2013; Yassin et al., 2017). They have become essential tools for management and decision making under uncertainty and non-stationarity, by providing the capability of prediction and support for scenario analysis regarding the quality and quantity of future Earth's resources (Kwakkel et al., 2016; Maier et al., 2016). These models continue to grow in complexity with our ever-growing understanding of underlying system processes, their heterogeneity, and feedback mechanisms (Razavi et al., 2012; Tetzlaff et al., 2008). This growth in complexity (and presumably model fidelity) has, however, resulted in large, computationally intensive models with many (sometimes hundreds of) "uncertain" parameters and factors whose effects on model behavior need to be characterized and understood (Kaizer et al., 2015; Oreskes, 2003; Razavi, 2017).

The pressing need to characterize how uncertainty in model parameters translates into uncertainty in model predictions has spawned development of a range of methods and tools for uncertainty analysis, rooted in probability theory. In most cases, these methods are based on the two traditional, forward- and inverse-problem approaches. The former propagates assumptions regarding uncertainties in system inputs or other properties (such as parameters and/or system structure) through the model to obtain some understanding regarding uncertainties in the model predictions *(e.g.,* Hong et al., 2006; Kunstmann et al., 2002). Conversely, the latter uses the information contained in the mismatch between model predictions and data to help identify "good" values for the model parameters, and to characterize their associated posterior uncertainty *(e.g.,* Beven and Binley, 1992; Smith and Marshall, 2008; Vrugt et al., 2003). A third, complementary approach that has gained momentum in recent years is one based on the paradigm of "sensitivity analysis" (SA), which seeks to illuminate the controls on model behavior, thereby characterizing the dominant controls on predictive uncertainty (Razavi and Gupta, 2015).

A fundamental basis for SA is an effect called the "Sparsity of

* Corresponding author. Global Institute for Water Security, University of Saskatchewan, Saskatoon, Saskatchewan, Canada.
E-mail address: saman.razavi@usask.ca (S. Razavi).

https://doi.org/10.1016/j.envsoft.2018.10.005
Received 18 May 2018; Received in revised form 4 October 2018; Accepted 23 October 2018
Available online 28 October 2018

Factors" principle. This principle, which originates from the field of Statistical Design of Experiments, states that the behavior of a process involving several variables is likely to be driven primarily by a small subset of these variables (Box and Meyer, 1986). In SA, where one aims to attribute the uncertainty in a model prediction to the uncertainties associated with different factors, we seek to answer the critical question: *when does uncertainty matter?* Imagine, for example, that SA could tell you the following about two parameters of a model: (i) parameter *A* is highly uncertain but does not strongly influence the uncertainty in model prediction, and (ii) parameter *B* is relatively certain while strongly influences (and can therefore cause substantial uncertainty in) the model prediction. It is clear that uncertainty with regards to parameter *B*, albeit small, is a dominant control of uncertainty in the problem at hand, whereas the large uncertainty in parameter *A* does not matter much. Such a characterization of uncertainty sources and their impacts is invaluable in guiding research towards reducing the uncertainties that matter, as it may point to the most important aspects of the problem at hand.

Early developments in SA were largely based on the notion of "local sensitivity", which derives point-based sensitivity measures specified locally around a nominal point in the problem space (Razavi and Gupta, 2015). This traditional practice, commonly referred to as "local sensitivity analysis" (LSA), is known to be incomplete and potentially misleading (Saltelli and Annoni, 2010), and has therefore evolved into a more advanced paradigm known as "global sensitivity analysis" (GSA) (Saltelli et al., 2008). GSA can be defined as a systems theoretic approach to characterizing the overall (average) sensitivity of one or more model responses across the factor space, by attributing the variability of those responses, to different controlling (but uncertain) factors (e.g., model parameters, forcings, and boundary and initial conditions). GSA, and SA in general, has a variety of applications, as outlined in Razavi and Gupta (2015), including:

- **Uncertainty Apportionment:** Attribution of total uncertainty in model responses to different factors (uncertainty sources) to identify where best to focus efforts for improved factor characterization so as to reduce the total uncertainty *(e.g.*, Chu-Agor et al., 2011).
- **Diagnostic Testing:** Assessment of similarities between the functioning of the model and the underlying real-world system, so as to assess the fidelity of the model structure, conceptualization, and parameterization *(e.g.,* Haghnegahdar et al., 2017).
- **Factor Prioritization and Function:** Identification of the factors that are more influential and contribute most significantly to variability and other characteristics of model response, and to understanding their role and function (*e.g.*, Muleta and Nicklow, 2005).
- **Factor and Model Reduction:** Identification of non-influential factors and/or insensitive (possibly redundant) components of model structure and parametrization so that they can be constrained or removed to simplify the model/analysis *(e.g.,* Touzani and Busby, 2014).

Here, we introduce a GSA toolbox, called VARS-TOOL, which includes a unique collection of state-of-the-art algorithms and tools for any of the applications outlined above, and beyond. The toolbox was designed to address the needs of any user (beginner to advanced) with any level of background knowledge of GSA and computer programming. It has been developed to improve upon existing software programs for GSA such as PSUADE (Gan et al., 2014), SAFE (Pianosi et al., 2015), DAKOTA (Adams et al., 2009), SimLab (JRC, 2008), and UQLab (Marelli and Sudret, 2014); Section 2 provides an overview of such improvements and the unique features of VARS-TOOL. Central to the development of VARS-TOOL is attention to ease of use and interpretability. While VARS-TOOL was originally developed using MATLAB, it has now also been written in C++ and built into the OSTRICH Software Toolkit that provides a model-independent interface for connecting VARS-TOOL with any simulation model (Matott, 2017). Being under continuous development, new capabilities, features and implementations in other computer programming languages are forthcoming. A well-designed Users' Manual provides detailed descriptions of the different functions within VARS-TOOL, along with relevant step-by-step examples. Our ultimate goal is to promote best practices in GSA applications within the Earth and environmental systems modelling community and beyond.

2. Why VARS-TOOL?

VARS-TOOL is a comprehensive, multi-approach, multi-algorithm toolbox equipped with a set of tools to enable GSA for any application, with a primary focus on dynamical Earth and environmental systems models. It is developed around the VARS (Variogram Analysis of Response Surfaces) theory and methodology, which provides a general framework that utilizes directional variogram and covariogram functions to characterize "global sensitivity", thereby providing a comprehensive set of global sensitivity indices with minimal computational cost (Razavi and Gupta, 2016a). VARS was developed to address two major challenges associated with GSA, as outlined in Razavi and Gupta (2015):

→ ***Ambiguous Definition of Global Sensitivity:*** different GSA methods are based in different philosophies and theoretical definitions of sensitivity, leading to different, even conflicting, assessments of the underlying sensitivities for a given problem.
→ ***Computational Cost:*** the cost of carrying out GSA can be large, even excessive, for high-dimensional problems and/or computationally intensive models, where cost (or "efficiency") is commonly assessed in terms of the number of required model runs.

The VARS approach can be seen as a "unifying theory" for GSA that places the different GSA theories and methods available in the literature on a common foundation. We say so because it re-defines GSA by characterizing a comprehensive spectrum of information about the underlying sensitivities of a response surface to its factors, including:

(1) local sensitivities, i.e., the partial derivatives of model responses with respect to different factors, and their global distributions across the factor space,
(2) the global distribution of model direct responses and the change in that distribution as a result of fixing one or groups of factors at different values within their uncertainty ranges, and
(3) the form and covariance structure of the response surface along the directions of different factors in the factor space.

The information types provided in points 1 and 2 above are, respectively, the bases for derivative-based methods such as elementary effects (Morris, 1991) and its extensions (Campolongo et al., 2007; Rakovec et al., 2014; Sobol' and Kucherenko, 2009) and direct-response-based methods such as Sobol' variance-based (Sobol, 2001), higher-order moment-based (skewness, kurtosis; Dell'Oca et al. (2017)), and general distribution form-based (Pianosi and Wagener, 2015) methods. VARS bridges these two philosophically different families of methods and further complements those with the information type provided in point 3 above; an information type that is unique to VARS.

Accordingly, VARS introduces a novel and general "variogram-based" paradigm for GSA that unifies and encompasses the pre-existing, widely used derivative-based (Morris, 1991) and variance-based (Sobol, 2001) approaches and their extensions as special/limiting cases. The theoretical relationship between VARS and the derivative/variance-based approaches, established in Razavi and Gupta (2016a), enables VARS to simultaneously generate both Morris and Sobol' sensitivity indices (including elementary effects and total-order effects) along with the recommended VARS-based "IVARS" indices, using a single

(common) set of sample points. In addition, VARS tackles the issue of "sensitivity" of the sensitivity analysis results to *perturbation scale* (scale issue in GSA) by providing sensitivity information spanning a range of scales across the factor space, from small-scale features such as roughness/noise to large-scale features such as multimodality (Haghnegahdar and Razavi, 2017; Razavi and Gupta, 2015).

A defining feature of VARS is its high efficiency and statistical robustness, enabling reliable and stable results with 1–2 orders of magnitude fewer sample points (model runs) than are required by alternative approaches, such as the derivative- and variance-based counterparts (Razavi and Gupta, 2016b). This computational efficiency is, in part, due to VARS being based on the information contained in *pairs* of points, rather than in individual points. As a result, VARS can be used to effectively and efficiently handle *high-dimensional problems*. This feature is important because, in practice, computational cost is a major reason why most applications of GSA (and of uncertainty analysis in general) have been limited to low-dimensional, simple (cheap-to-run) models. This is also related to the curse of dimensionality in which, as the problem dimension (e.g., number of parameters) grows, the volume of the problem space increases so rapidly (exponentially) that the available sample density becomes too sparse to be able to properly characterize the problem space; as such, the size of the sample required (i.e., number of model runs) for a stable, robust, and statistically sound assessment typically grows exponentially with dimension. The recent survey by Sheikholeslami et al. (2018) finds that ~70 percent of GSA applications in the environmental modelling literature have focused on models with less than 20 parameters, which falls well below the numbers of factors that comprise complex, state-of-the-art models. This may be seen as *paradoxical* to the underlying goal of GSA, which is to facilitate understanding of the behaviors of complex models, nowadays involving tens to hundreds (or more) of factors.

In addition to providing the first software implementation of the VARS methodology, VARS-TOOL includes a range of other algorithms and software tools, as listed in Table 1 and described below (the list is continually being updated). Notably, VARS-TOOL includes:

(a) A suite of GSA algorithms in addition to VARS. This feature enables the users to use an algorithm that they are most comfortable with and/or to compare the performance and results of different algorithms together and with those provided by VARS (Razavi and Gupta, 2016a, 2016b).
(b) A suite of sampling strategies. A particularly important feature is the implementation of Progressive Latin Hypercube Sampling (PLHS), which enables the progressive generation of additional sample points in the factor space, while continuing to preserve the distributional properties of interest (Sheikholeslami and Razavi, 2017).
(c) Novel tools for assessing the time-varying nature of sensitivities of dynamic systems models. This capability enables the user to compute "time-varying" and "time-aggregate" sensitivity indices of model state and output variables through the Generalized Global Sensitivity Matrix (GGSM) approach (Gupta and Razavi, 2018; Razavi and Gupta, in review).
(d) A strategy for "factor grouping" based on their importance and function. This feature enables classification of factors into groups, which greatly facilitates the analysis of problems that are of very high-dimension (having large numbers of factors (Sheikholeslami et al., 2018)).
(e) A measure of robustness and convergence of sensitivity analysis. Via statistical bootstrapping, this feature enables estimation of confidence intervals on sensitivity indices as well as the reliability, robustness, and convergence of factor ranking or grouping over time throughout a GSA experiment (Razavi and Gupta, 2016b; Sheikholeslami et al., 2018).
(f) Model emulation strategies to facilitate "model crash handling". This feature maintains the reliability and robustness of a GSA experiment when the simulation model used fails to return response values for some sample points in the factor space (Sheikholeslami et al., in prep.).

Importantly, VARS-TOOL comes with a visualization tool, that can work in online mode to enable the user to monitor the real-time performance and evolution of a GSA experiment, and to process its intermediate results while assessing stability and convergence. Finally, the inclusion of several test functions and real-world case studies, including the HBV-SASK rainfall-runoff model, enables the use of VARS-TOOL for a range of learning, teaching, and research purposes. By incorporating the aforementioned diversity of tools and features within a *single* platform, VARS-TOOL is intended to conveniently provide the user with the ingredients necessary for conducting exploratory research with a view to discovering new directions for advancing the field of sensitivity and uncertainty analysis. In the remainder of this paper, we provide details on the features and tools outlined above.

3. Sensitivity analysis algorithms

VARS-TOOL includes a range of well-known "model-free" (or "model-independent") GSA algorithms that work with any model of any degree of complexity. These algorithms are based on derivative- and variance-based approaches that are commonly used and reported in the literature, and the recently developed, more general variogram-based approach that bridges across the aforementioned two approaches. A brief description of the algorithms is given below, and details regarding their numerical implementation (sampling strategies and computational costs) are provided in Section 5.

The derivative-based approach: This approach is a natural extension to local sensitivity analysis wherein the partial derivatives of a model response with respect to different model inputs at one *base point* are computed numerically and interpreted as indices of local sensitivity. The associated algorithm (see Table 1) generates "globally aggregated measures of local sensitivities" by computing the partial derivatives (called "elementary effects" by Morris, 1991) at many sample points and combining them together in "some" way (e.g., by taking the mean and/or standard deviation) to generate indices for global sensitivity. Each index is different in the way that it characterizes the distributional properties of partial derivatives (for details, see Razavi and Gupta (2015)). The VARS-TOOL function "main_Morris.m" implements the algorithm and computes different variations of elementary effects-based sensitivity indices proposed by Morris (1991), Campolongo et al. (2007), and Sobol' and Kucherenko (2009). The user is required to select a step size for numerical approximation of the partial derivatives, which is typically recommended (arbitrarily) to be 1–10% of the input range.

The variance-based approach: This approach is based on analysis and decomposition of the variance of model response, so as to interpret the contribution of different factors in explaining this variance as an index for global sensitivity. The VARS-TOOL function "main_Sobol.m" (Fig. 1) implements the algorithm of Saltelli et al. (2008) to decompose the total variance of model response (when all the factors are varied within user-selected ranges) into its components arising from individual inputs and their interactions. Algorithm outputs include the "main effects" associated with individual factors (discarding the role of possible interactions with any other factors), and the "total-order effects" that combine the contribution of individual factors and all higher-order interactions with other factors. Typically, the total-order effect is considered to be an effective index for global factor sensitivity.

The variogram-based approach: This approach is based on Variogram Analysis of Response Surfaces (VARS) that provides a general, comprehensive framework that unifies and extends upon both the derivative- and variance-based approaches. VARS was specifically developed to address a major weakness of the aforementioned approaches, which is that neither considers or accounts for the *spatially*

Table 1
A summary of tools, capabilities, and test beds available in VARS-TOOL.

Sensitivity Analysis Algorithms	Sampling Strategies	Other Features	Test Functions	Real-World Case Studies
Derivative-based (Morris) Algorithm that computes "elementary effects" based indices of Morris (1991), Campolongo et al. (2007), and Sobol and Kucherenko (2009) (main_Morris.m)	Random Sampling (RND)	Bootstrapping to estimate confidence intervals on sensitivity indices and robustness of factor rankings (Razavi and Gupta, 2016b)	These functions provide synthetic examples for systems with scalar-model outputs:	The HBV-SASK hydrologic model set up for two case studies, the Bow and Oldman Rivers. These case studies provide real-world examples of dynamical systems with time-dependent (vector) outputs
	Latin Hypercube Sampling[a] (McKay et al., 1979) (LHS)		Multi-scale wavy function Dimensions = 6 (Razavi and Gupta, 2016a)	
Variance-based (Sobol) Algorithm that computes "main effect" and "total-order effect" indices (Saltelli et al., 2008) (main_Sobol.m)	Symmetric Latin Hypercube Sampling (Ye et al., 2000) (SymLHS)	Factor grouping to identify and group factors of comparable effects (Sheikholeslami et al., 2018)	Ishigami Dimensions = 3 (Ishigami and Homma, 1990)	The MESH (Modélisation Environmentale communautaire – Surface & Hydrology) land surface-hydrology model set up for Canadian watersheds. MESH is based on Canadian Land Surface Scheme (CLASS). This is a computationally-intensive, physically-based Earth system model connected to VARS-TOOL via OSTRICH Toolkit
Variogram-based Algorithm that computes IVARS indices (including "total-variogram effects"), "total-order effects", and "elementary effects" all together, through STAR-VARS (Razavi and Gupta, 2016b) (main_VARS.m)	Progressive Latin Hypercube Sampling (Sheikholeslami and Razavi, 2017) (PLHS)	Model emulation to handle failures of simulation models (Sheikholeslami et al., in prep.) Online visualization to monitor stability and convergence of GSA experiments	g-function (SobolG) Dimensions = D (Saltelli et al., 2008)	
The Generalized Global Sensitivity Matrix (GGSM) Algorithm coupled with STAR-VARS that computes "time-varying" and "total-period time aggregate" sensitivity indices (Gupta and Razavi, 2018; Razavi and Gupta, in review) (main_VARS.m)	Halton Sequence[a] (Halton, 1960) (Halton) Sobol Sequence[a] (Sobol', 1967) (Sobol)	Interface to link VARS-TOOL with any model in any programing language and operating system	Sobol-Levitan (soblev99) Dimensions = D (Sobol and Levitan, 1999) Welch function Dimensions = 20 (Welch et al., 1992)	
Monte-Carlo Filtering Algorithm (also called Regional Sensitivity Analysis) (Spear et al., 1994) (main_RSA.m)	Star-based Sampling that links with any of the above (Razavi and Gupta, 2016b) (STAR)		Levy function Dimensions = 2 (Laguna and Martí, 2005) Eggholder Dimensions = 2 (Jamil and Yang, 2013) Langermann function Dimensions = D (Jamil and Yang, 2013)	

[a] In the MATLAB version, these sampling strategies are built-in MATLAB functions that are available to VARS-TOOL.

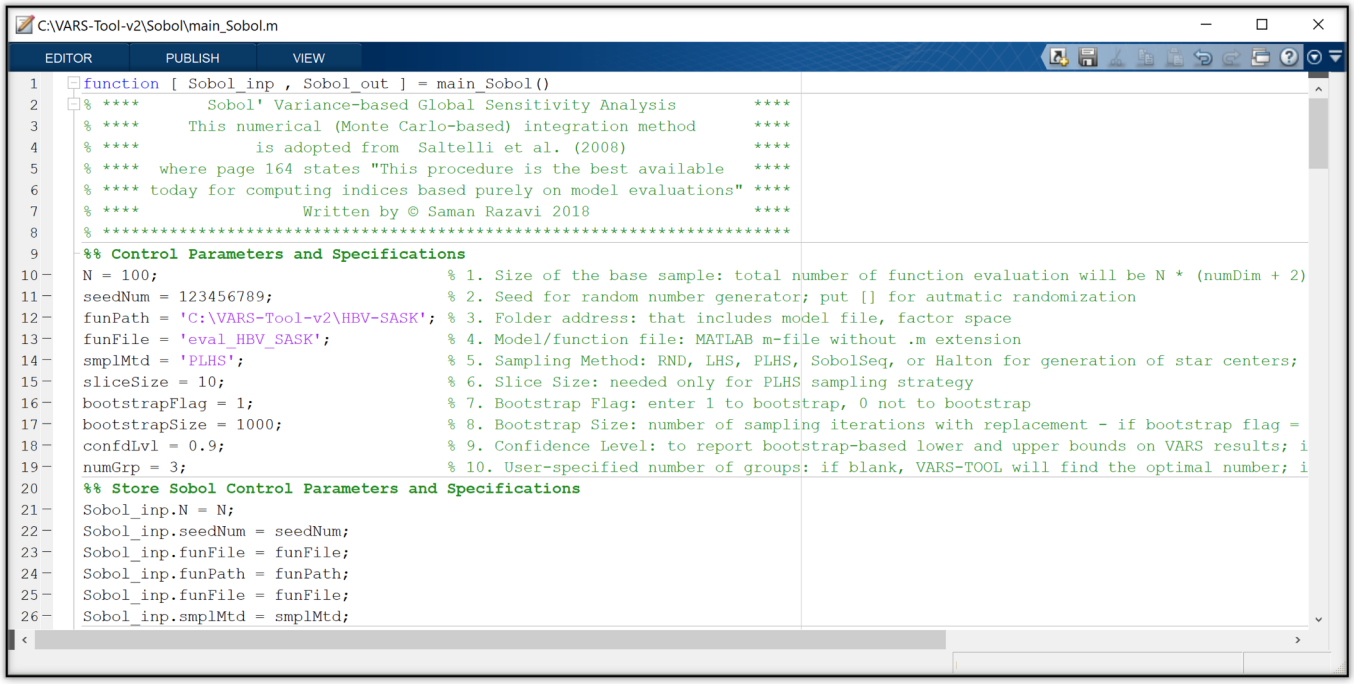

Fig. 1. A screenshot of main_Sobol.m function in VARS-TOOL.

ordered structure of the model response in the factor space. In other words, they ignore the fact that the response values are *not* randomly distributed throughout the factor space. Conversely, VARS recognizes that there is a *spatially continuous covariance structure* to the model response, and hence also to its partial derivatives. To extract and analyze this structural information, VARS uses *anisotropic* variogram and covariogram functions of the model response to generate "directional variograms" associated with each of the model factors as a basis for a comprehensive characterization of global sensitivity.

The directional variogram represents the variance (i.e., rate of variability) of the response caused by perturbing that factor across a full range of "perturbation scales", while all other factors are also varied in the factor space. Fig. 2 shows schematic directional variogram and covariogram functions and how they relate to derivative- and variance-based approaches. By definition, the left end point of a variogram (where the perturbation scale, h, is small) represents derivative information of the underlying process, while its right end point (where h is large) represents the process variance; the variance-based total-order effect can be estimated via mathematical manipulation of the variogram and covariogram functions. For the detailed analytical relationships between the variogram-, derivative-, and variance-based theories, see Razavi and Gupta (2016a). The sensitivity indices in VARS are computed by integrating the directional variograms, resulting in a comprehensive set of indices for global sensitivity called "Integrated Variograms Across a Range of Scales" or IVARS. The $IVARS_{50}$ index, also referred to as the "total-variogram effect", that integrates the variogram across the full range of perturbation scales is the most comprehensive variogram-based index for global sensitivity.

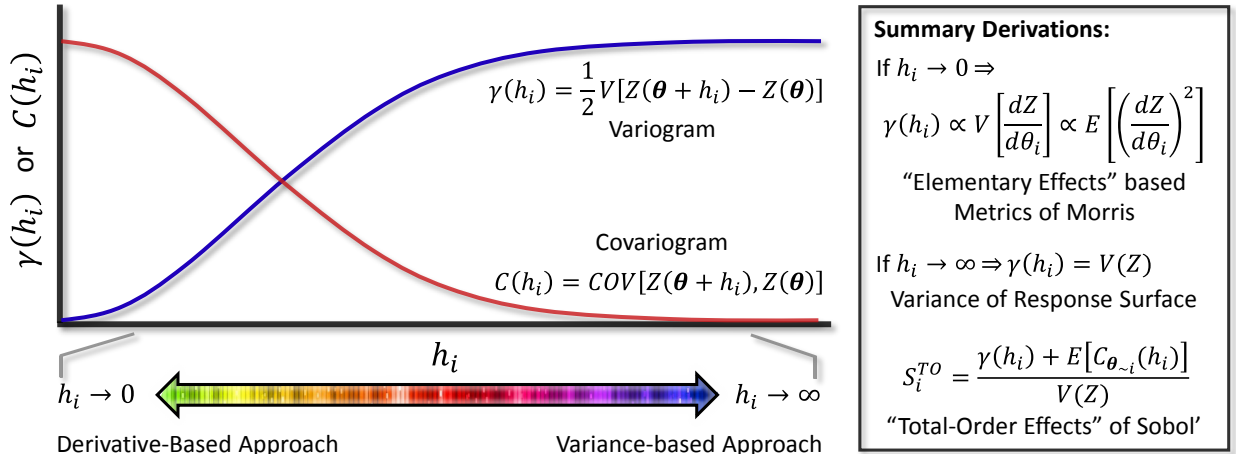

Fig. 2. Using the directional (anisotropic) variogram (γ) and covariogram (C) concepts, the variogram-based approach provides a general, unifying theory that bridges across the derivative- and variance-based approaches. θ is the set of factors, θ_i is the i^{th} factor, and $\theta_{\sim i}$ is the set of all factors excluding the i^{th} factor; Z and h_i represent model response and perturbation scale, respectively. Note that the relationship between the total-order effect and the variogram and covariogram functions holds for any h_i, under the constant mean assumption; this assumption tends to be more accurate for smaller h_i, and therefore, VARS estimates S_i^{TO} (also referred to as VARS-TO) based on the smallest numerical h_i.

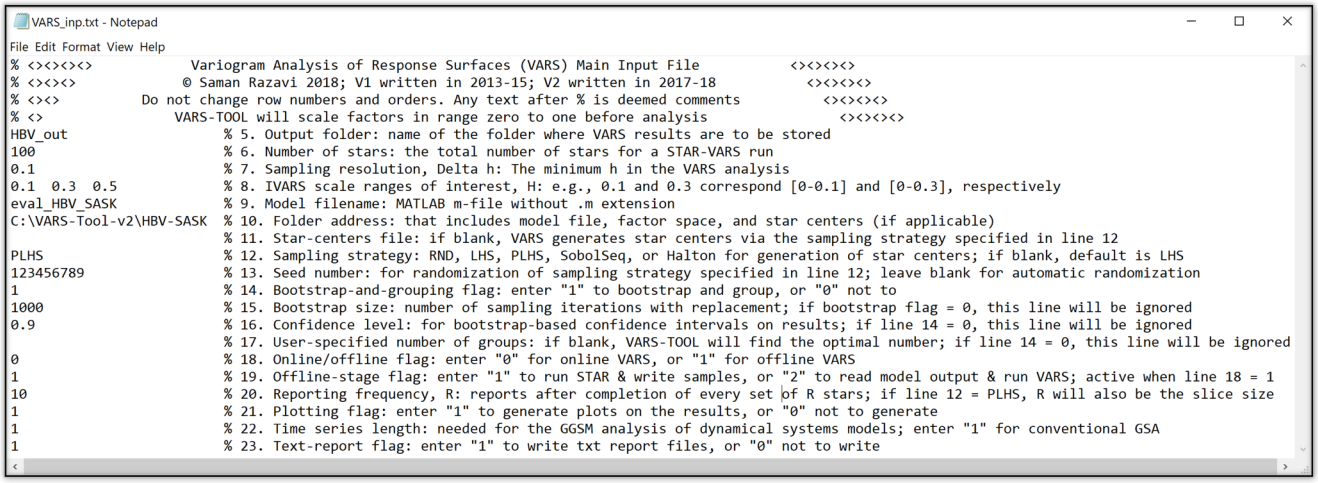

Fig. 3. A screenshot of VARS_inp.txt (main input file), where the algorithm parameters, the files containing the model to run and factor ranges, etc. are specified. The percent sign (%) at each line denotes that the following information are comments and will not be used by VARS. These specifications can also be directly provided in the source codes.

A specific implementation of VARS, called STAR-VARS, is incorporated within VARS-TOOL. This implementation utilizes a form of star-based sampling (called STAR), developed by Razavi and Gupta (2016b). The VARS-TOOL function "main_VARS.m" computes the IVARS indices for global sensitivity along with the aforementioned derivative- and variance-based indices, all together in a single run that utilizes the same sample points. Therefore, this function is a multi-method platform that provides the user with a sensitivity assessment that simultaneously includes the derivative-, variance-, and variogram-based indices. Fig. 3 shows an example input text file to this function, where the user can provide the specifications of model, factor ranges, algorithm parameters, etc. The text-file-based interface with this function was designed for ease of use.

To be comprehensive, VARS-TOOL also includes an implementation of the so-called "Regional Sensitivity Analysis" (RSA) approach, that is essentially Monte-Carlo filtering (Hamby, 1994; Spear et al., 1994). The heuristic RSA approach is commonly used to partition the marginal distribution of sample points obtained for each factor into two (or more) distributions based on empirically selected threshold values for model response. The idea is that if the factor does not have a significant impact on model response throughout the factor space, the two distributions should be statistically indistinguishable. The VARS-TOOL function "main_RSA.m" implements this approach and utilizes the Smirnov test to quantify the extent to which the two distributions are different, thereby providing indices for global sensitivity.

Other classic "model-based" GSA approaches such as factorial design-, correlation-, and regression-based approaches are not included in VARS-TOOL, because of their limited utility in the design and analysis of computer-based simulation experiments. These approaches are generally unsuitable for sensitivity analysis of complex dynamical systems models as they are based in the a priori assumption of a particular mathematical form (typically linear or polynomial) for the underlying model response surface. In addition, these methods are particularly prone to curse of dimensionality (see discussion in Razavi and Gupta (2015)).

4. Sensitivity analysis of dynamical systems models

Most approaches to global sensitivity analysis (GSA) do not adequately account for the *dynamical nature* of Earth and environmental systems models; the fact that such models, in real-world applications, produce dynamical time-evolving responses to dynamical time-evolving perturbations/inputs. Gupta and Razavi (2018) highlighted this fact, revisited the fundamental basis of GSA for dynamical systems models, and developed a sensitivity analysis framework from first principles based on computation of a "Global Sensitivity Matrix" (GSM) that quantifies the sensitivity information contained in trajectories of partial derivatives of the dynamical model responses with respect to controlling factors. Razavi and Gupta (in review) extended and generalized this approach to accommodate (a) any GSA philosophy including derivative-, variance-, and variogram-based approaches and (b) any model response of any type including time series of model state or output variables and their transformations; this approach is referred to as "Generalized Global Sensitivity Matrix" (GGSM).

VARS-TOOL includes an efficient implementation of the GGSM approach coupled with STAR-VARS that enables a multi-method GSA of dynamical systems models that accounts for the temporal dynamics of such models while enabling efficient comparison of the results provided by philosophically different approaches (e.g., derivative-based versus variance-based). This implementation generates the following types of indices of global sensitivity:

(a) **"Time-varying" sensitivity indices:** These indices are in the form of time series that reveal the time-dependent sensitivities of model responses (over the course of a simulation time period) to its controlling factors.
(b) **"Time-aggregate" sensitivity indices:** These indices are in the form of summary statistics that aggregate and summarize the dynamical sensitivity information of model responses over one or multiple time periods of interest.

The time-varying indices enable the user to better understand the model/system behavior over time, in response to system forcings/boundary conditions. This capability facilitates: (i) diagnostic testing and detection of potential defects in different parts of a model, thereby helping to improve model realism, and (ii) attribution of variability and therefore uncertainty in the model responses to different factors (model parameters, forcings, boundary conditions, etc.), thereby helping to pinpoint the dominant controls of predictive uncertainty at different points in time. When summary information regarding the time-varying sensitivities is required, the time-aggregate indices become more useful.

A second defining feature of the GGSM approach is that it does not require that observed data on system responses be available, and so can be used to assess the internal functioning of a model and the controls exerted by different factors of any of its components, including both states and fluxes. This is important because (1) our interest when

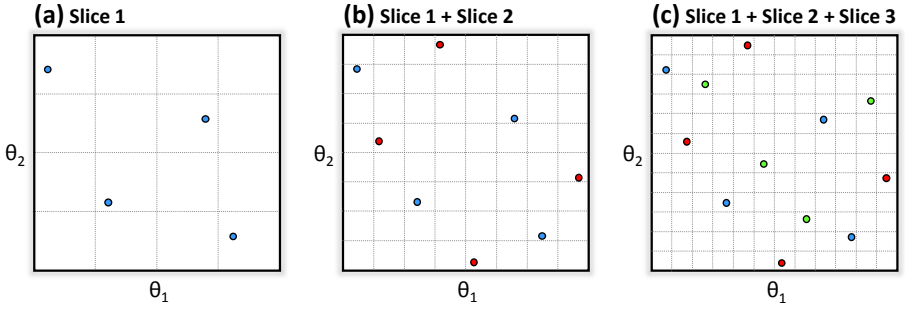

Fig. 4. Illustration of Progressive Latin Hypercube Sampling (PLHS) on a 2-dimensional factor space with 3 slices and a slice size of 4. (a) Sample points of slice 1, ensuring every one-fourth of space at each direction receives one point. (b) Sample points of slice 2 added to those of slice 1, ensuring every one-eighth of space at each direction receives one point. (b) Sample points of slice 3 added to those of slice 1 and slice 2, ensuring every one-twelfth of space at each direction receives one point. Points of slices 1, 2, and 3 are in blue, red, and green, respectively.

conducting a GSA is not limited to response variables for which observations are available, and (2) the sensitivity assessments cannot be obscured or distorted by errors and uncertainties that unavoidably exist in observed response data (Gupta and Razavi, 2018).

The VARS-TOOL function "main_VARS.m" is enabled with the GGSM approach, which can be activated by a switch on line 22 of VARS_inp.txt (Fig. 3) or within the source code. Upon activation, VARS-TOOL can generate, within a single run (i.e., a single sample set), time-varying and time-aggregate GSA indices based on the derivative-based (e.g., Morris elementary effects), variance-based (e.g., Sobol' total-order effects), and variogram-based (e.g., VARS total-variogram effects) approaches. This capability enables a user to explore, compare, and contrast the assessments provided by these three approaches to GSA.

5. Sampling strategies

Sampling strategies are necessary and fundamental components of any algorithm for sensitivity and uncertainty analysis of computer simulation models. VARS-TOOL includes a variety of sampling strategies, as listed in Table 1. The derivative-, variance-, and variogram-based GSA algorithms outlined above all follow a *two-level* sampling procedure in the factor space. In the first level, they can employ any of the general-purpose, *randomized sampling strategies* listed in Table 1 to select a set of N "base points" randomly distributed uniformly throughout the factor space. In the second level, these points are used as starting points for the algorithm-specific, *structured sampling strategies* used to acquire the remaining sample points required for calculating the algorithm-specific sensitivity indices. VARS-TOOL includes both "single-stage" and "progressive" (also called "sequential" or "multi-stage") strategies for random sampling.

Because single-stage sampling strategies such as Latin hypercube sampling (LHS; McKay et al., 1979) and symmetric Latin hypercube sampling (SLHS; Ye et al., 2000) generate an entire sample, consisting of a pre-specified number of points, all at once, such strategies are effective only if the user is fairly confident about the proper sample size before beginning the analysis, as any subsequent enlargement of the sample size will generally fail to preserve the distributional properties of interest. In contrast, progressive sampling strategies enable sequential (or multi-stage) generation of the sample points. Examples include the traditional Halton and Sobol sequences (also called low-discrepancy sequences), and the novel method of progressive Latin hypercube sampling (PLHS) recently developed by Sheikholeslami and Razavi (2017). Note that VARS-TOOL provides the first available software implementation of PLHS.

PLHS sequentially generates sample points while progressively preserving important distributional properties of interest (Latin hypercube properties, space-filling, etc.), as the sample size grows. Fig. 4 illustrates how PLHS successively generates a series of smaller subsamples (slices) such that (1) the first slice is Latin hypercube, (2) the progressive union of slices remains Latin hypercube and achieves maximum stratification in any one-dimensional projection, and as such (3) the entire sample is Latin hypercube. PLHS has been shown to be a superior strategy that scales effectively with the size and dimensionality of the problem under investigation (Sheikholeslami and Razavi, 2017).

Once the set consisting of N base points is generated randomly by any of the sampling strategies mentioned above, it is used by the structured sampling strategies as starting points for the generation of the remaining sample points. In the derivative-based approach (Campolongo et al., 2007; Morris, 1991), each of the N base points serves as the starting point for a chain of points constructed by changing one factor at a time (by a given step size), resulting in a total number of N(D + 1) sample points, where D is the number of factors. In the variance-based approach (Saltelli et al., 2008), structured sampling is conducted via matrix manipulation of the base sample matrix (consisting of the N base points), resulting in a total number of N(D + 2) sample points. In the variogram-based approach, and particularly in the STAR-VARS implementation, the N base points are used as "*star centers*" for STAR sampling (Razavi and Gupta, (2016b)). In this approach, equally spaced points are sampled (Δh apart) around each star center along every dimension of the factor space (Fig. 5), resulting in a total number of N(D(1/Δh-1) + 1) sample points. The distance Δh, referred to as the "*resolution*" of sampling, is specified by the user; a Δh of 0.1 of the factor range is recommended, while smaller values can be selected to yield more accurate results.

6. Dealing with high-dimensional problems: factor grouping

The challenge of conducting a GSA, monitoring its convergence and robustness, and interpreting the associated results can easily become non-trivial when the problem has more than ~20–30 factors.

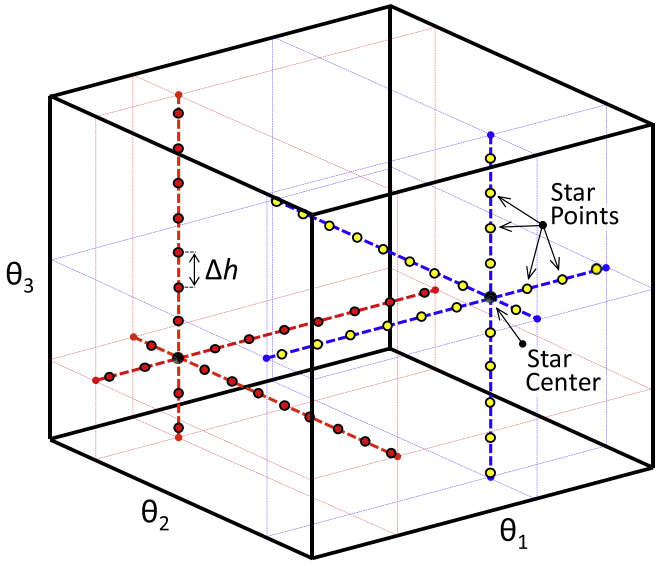

Fig. 5. Illustration of the STAR sampling strategy with N = 2 (number of star centers or base points), D = 3 (number of factors), and Δh = 0.1 of factor range.

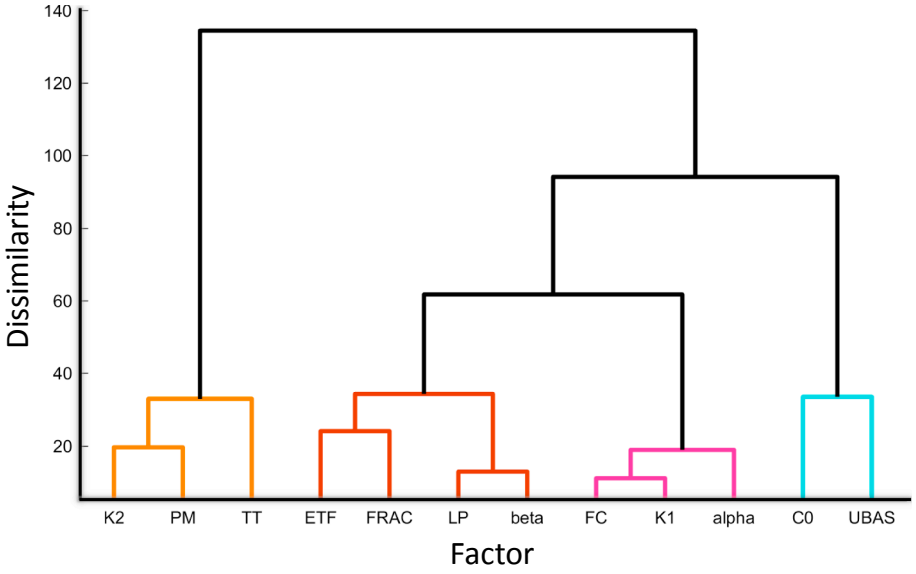

Fig. 6. An example dendrogram for factor grouping generated by VARS-TOOL for the HBV-SASK hydrologic model on the Banff River Basin (see Section 9). This grouping is based on the sensitivity of mean-squared-errors performance metric on log-transformed simulated and observed streamflows over the historical record to the 12 model parameters. The parameters are sorted from the most influential (to the left) to the least influential (to the right), and colored groups of parameters correspond to the optimal grouping obtained by the elbow method.

Therefore, VARS-TOOL includes an innovative "factor grouping" strategy that employs a clustering mechanism developed by Sheikholeslami et al. (2018) to handle high-dimensional problems, involving tens to hundreds of factors. This strategy, integrated together with bootstrapping (see Section 7.1), monitors GSA performance and clusters factors into groups of similar properties based on their sensitivity and function. The resulting groups can be of any size, depending on the problem at hand.

For example, in a problem with 60 factors, VARS-TOOL can be used to cluster the factors into 5 groups of factors termed "strongly influential", "influential", "moderately influential", "weakly influential", and "non-influential". The number of groups of interest can be directly specified by the user in VARS_inp.txt; as a convenient feature, however, if the user does not specify a preferred number of groups, VARS-TOOL will suggest an "optimal" number of groups, based on maximizing the distinctions between the groups through the so-called "elbow method" (Sheikholeslami et al., 2018). Fig. 6 shows an example dendrogram generated by VARS-TOOL and the associated optimal grouping.

This grouping capability is particularly beneficial when dealing with high-dimensional problems, where the user typically does not need to obtain an exact ranking for the many factors. For example, in a problem with 100 factors, it may not matter whether a factor is the 30th or the 31st in terms of importance, and obtaining a robust answer to this question may require running the model an excessively large number of times. Instead, by assigning factors having similar importance into groups, VARS-TOOL provides a means for making the problem more tractable within a limited number of model runs.

This capability scales the computational demand of a GSA algorithm with the users' need in terms of precision in factor rankings, and as such, results in significant improvements in stability and rate of convergence.

7. Other important features of VARS-TOOL

7.1. Bootstrapping: Characterizing confidence and robustness

In any analysis based on statistical sampling, it is essential to be able to assess the degree of confidence one can place in the results, so as to get a sense of reliability, robustness, and sufficiency of the analysis. In VARS-TOOL, the method of statistical bootstrapping is used to (1) infer confidence intervals around the estimates of sensitivity indices at any user-specified confidence level, and (2) provide an assessment of the "robustness" that can be associated with the factor rankings. While the use of bootstrapping for the former is common in GSA literature, in the latter, the definition and implementation of robustness (sometimes termed "reliability" as well) of a factor ranking, developed by Razavi and Gupta (2016b) and further extended by Sheikholeslami et al. (2018) for groups of factors, is unique to VARS-TOOL.

Consider, for example, a problem with five factors. After 200 model runs (sample size = 200), one may find that the sensitivity index associated with factor 3 has been estimated to have a value of 0.4, but with a confidence interval of [0.3–0.45] at the 90% confidence level. Meanwhile, its factor ranking may be estimated as 2 (meaning that it is the second most influential of the 5 factors) with an assessed robustness of 70% (meaning that factor 3 is exactly rank 2 with 70% probability). With larger sample sizes, we can expect that confidence intervals will become progressively narrower and the robustness estimates progressively improve, thereby increasing our confidence in the results of the statistical analysis. Note that the reason that we will typically not have 100% confidence in the results is that the analysis can be affected by "sampling variability" in the generation of the sample points used, and so can provide somewhat different results if the experiment is re-run with a different (but equally representative) set of samples.

However, increasing the sample size can help to improve the reliability and robustness of the analysis. So, for example, increasing the sample size to 500 in the example above could alter the sensitivity index of factor 3 to have a value of 0.38 with a 90% confidence interval of [0.35–0.41], and a factor ranking of 2 but with an increased robustness of 90%. In this way, statistical bootstrapping provides a way of assessing the robustness, stability, and convergence of a GSA experiment, with narrower 90% (or 95% etc.) confidence intervals and higher factor ranking robustness indicating greater reliability of the results.

As a caveat, note that statistical bootstrapping (Efron, 1992) is a procedure based in sampling with replacement from the existing set of sample points to obtain estimates of the underlying statistical properties of a sample, and so its validity depends on the assumption that the existing sample is a sufficiently representative sample of the underlying population (Beran, 1997; Davison et al., 2003). In practice, this means that the sample size used for the analysis must be large enough to adequately span the factor space of interest. One way to assess this is to repeat the analysis with progressively larger sample sizes, which is easily facilitated by the PLHS method of progressive sampling discussed in Section 5.

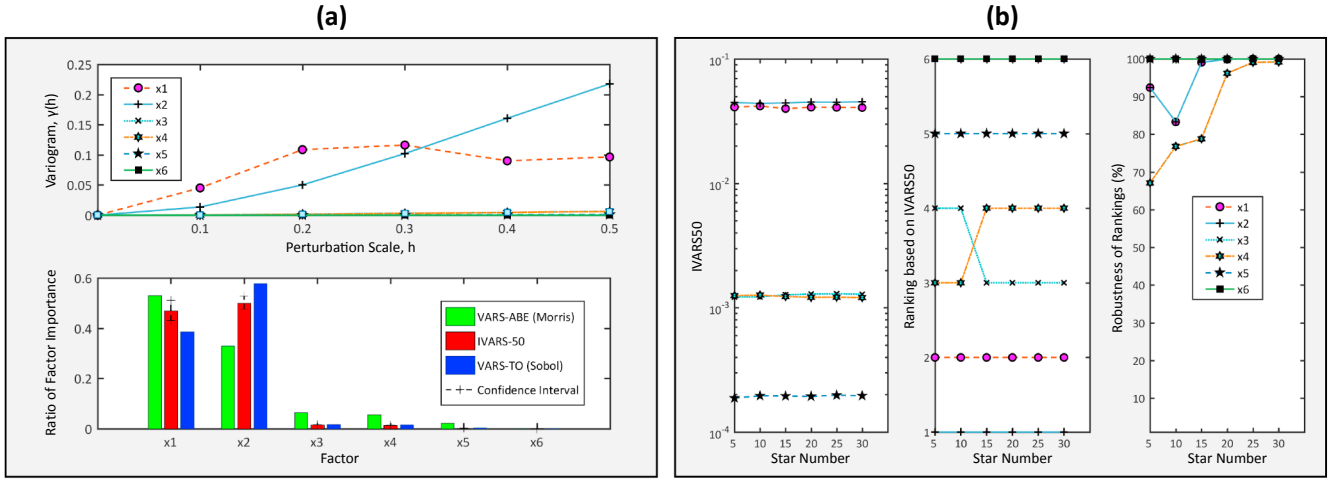

Fig. 7. Visualization of GSA results in VARS-TOOL. Plot (a) shows detailed GSA results, such as directional variograms that manifest sensitivity across the range of perturbation scales and bar charts that indicate relative importance of different factors according to the derivative-, variance-, and variogram-based approaches. Plot (b) shows the evolution and convergence of sensitivity indices and rankings during the execution of a GSA experiment.

7.2. Reporting and visualization: Monitoring stability and convergence

VARS-TOOL provides an "on-line" visualization and reporting capability that enables monitoring the execution of GSA experiments and evaluation of their intermediate/progressive performance. Fig. 7 shows example plots generated using VARS-TOOL for the Razavi-Gupta's "6D multi-scale wavy function", developed in Razavi and Gupta (2016a), in on-line mode. These plots enable the user to monitor how the estimates of factor sensitivities and rankings may be changing with increasing sample size (as more model evaluations become available).

Along with comprehensive, frequently generated report files (provided in text format) in the course of a GSA experiment, these plots provide the capability to visually monitor robustness, stability, and convergence, and terminate an analysis when required. Furthermore, this capability in conjunction with a sequential sampling strategy (e.g., PLHS) can resolve the problem that a user will typically not know a priori what a suitable sample size may be for a GSA experiment. As a result, the GSA can run only for an *optimal* number of model runs, thereby maximizing computational efficiency.

7.3. Model emulation: Handling model crashes

An issue that is known to arise when running some complex Earth and environmental systems models is that they may fail, while running, under particular factor configurations and values. Such failures (often called "*crashes*") can, for example, be caused by specifying unrealistic combinations for the parameter values that violate model assumptions, or by numerical instabilities that arise under some combinations of parameters and forcings.

In practice, if the occurrence of such model failures is not properly dealt with, the entire GSA experiment can fail. Most implementations of GSA algorithms require that meaningful model responses be provided for every sample point generated. However, execution of the model code may fail to return a response if it crashes. When this happens, modellers often end up re-doing the entire analysis with more conservative (tighter) factor ranges, resulting in extra computational cost (model runs) and also changing the definition of the original problem.

To address this issue effectively and efficiently, VARS-TOOL is enabled with a "crash-handling module" that is designed to complete the GSA while minimizing the impact of model crashes on the analysis. For this purpose, VARS-TOOL employs model emulation techniques (Razavi et al., 2012) to generate surrogate values of model outputs when a model fails. Theoretical details of this module are available in Sheikholeslami et al. (in prep.).

7.4. Running with models implemented in any programming language and operating system

VARS-TOOL programs, both the MATLAB version and C++ version implemented within OSTRICH, have the capability to work with any model implemented in any programming language and running on any operating system. VARS-TOOL has two modes for execution:

- **On-line (internal):** This mode can be used when the computer simulation model is set up to be called and run through the MATLAB or OSTRICH environments.
- **Off-line (external):** This mode is to be used when the computer simulation model needs to be run externally, outside the MATLAB or OSTRICH environments.

In the off-line mode, the model simulations can be performed independently of the VARS-TOOL software programs. In this mode, (1) VARS-TOOL first generates the locations for all the sample points required for the analysis and stores them in a text file. (2) Then the user should conduct model simulations externally for all of the sampled points and store the respective model response values (e.g., in a text file). These runs can be parallelized if multiple processors are available. (3) VARS-TOOL reads in the model runs results, conducts the analysis, and generates sensitivity indices. Importantly, the VARS-TOOL version within OSTRICH supports Message Passing Interface (MPI)-based parallel processing on both Windows and Linux systems, which can significantly increase computational efficiency of a GSA experiment by parallelizing model runs.

8. Test functions

Test functions are useful for learning, understanding, testing, and benchmarking the performance of different GSA algorithms. They have the advantage that (1) they are computationally inexpensive and fast-to-run and (2) their characteristics are known a priori, and in most cases, controllable by the user.

VARS-TOOL embeds a diverse set of test functions, demonstrating a wide range of possible characteristics such as nonlinearity, unimodality and multi-modality, irregularity, roughness and smoothness,

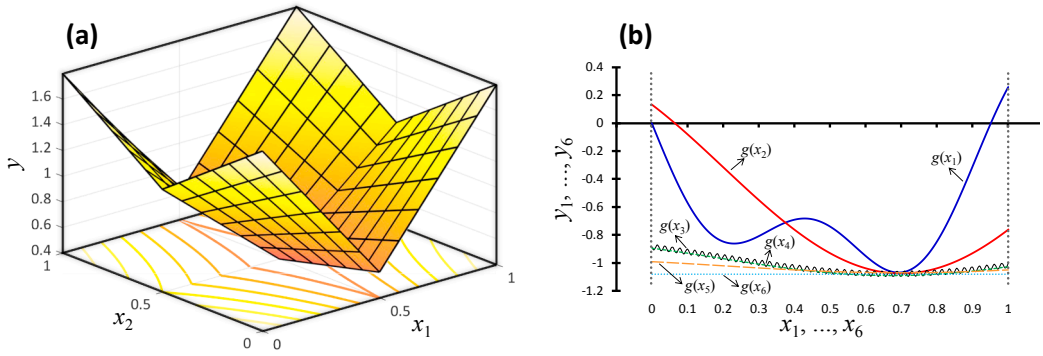

Fig. 8. The artificial response surfaces of test functions: (a) g-function and (b) 6D multi-scale wavy function. g-function demonstrates strong interaction effects and can be of any number of dimensions but the general structure and characteristics of the surface along the different dimensions are the same. 6D multi-scale wavy function is the additive effect of 6 one-dimensional functions, each demonstrating a distinct structure and characteristics, from small-scale features such as roughness to large-scale features such as multi-modality.

interaction, and high-dimensionality. These test functions provide "artificial" response surfaces (also called landscapes) that may share some characteristics with the response surfaces of real-world modelling problems; for a discussion of the latter see Duan et al. (1992).

In addition to including test functions designed originally for research on sensitivity and uncertainty analysis, VARS-TOOL includes functions popular in the field of optimization research. The test functions are listed in Table 1, and the response surfaces for two of them (G-function and 6D multi-scale wavy function) are shown in Fig. 8. When choosing a test function for a particular analysis, users might first consider which test function presents similar structure and characteristics to those of the underlying response surface of the problem at hand. This question is non-trivial, but Razavi and Gupta (2015) provide some insight.

9. Hydrologic case studies

VARS-TOOL includes software of a hydrologic model called HBV-SASK, which is an interpretation of Hydrologiska Byråns Vattenbalansavdelning model (Lindström et al., 1997), developed by the first author for educational and research purposes. Fig. 9 shows the architecture of HBV-SASK and its inputs, process parameters, and equations. In addition to those process parametrizations, daily potential evapotranspiration (PET) is computed using equation PET = (1 + ETF

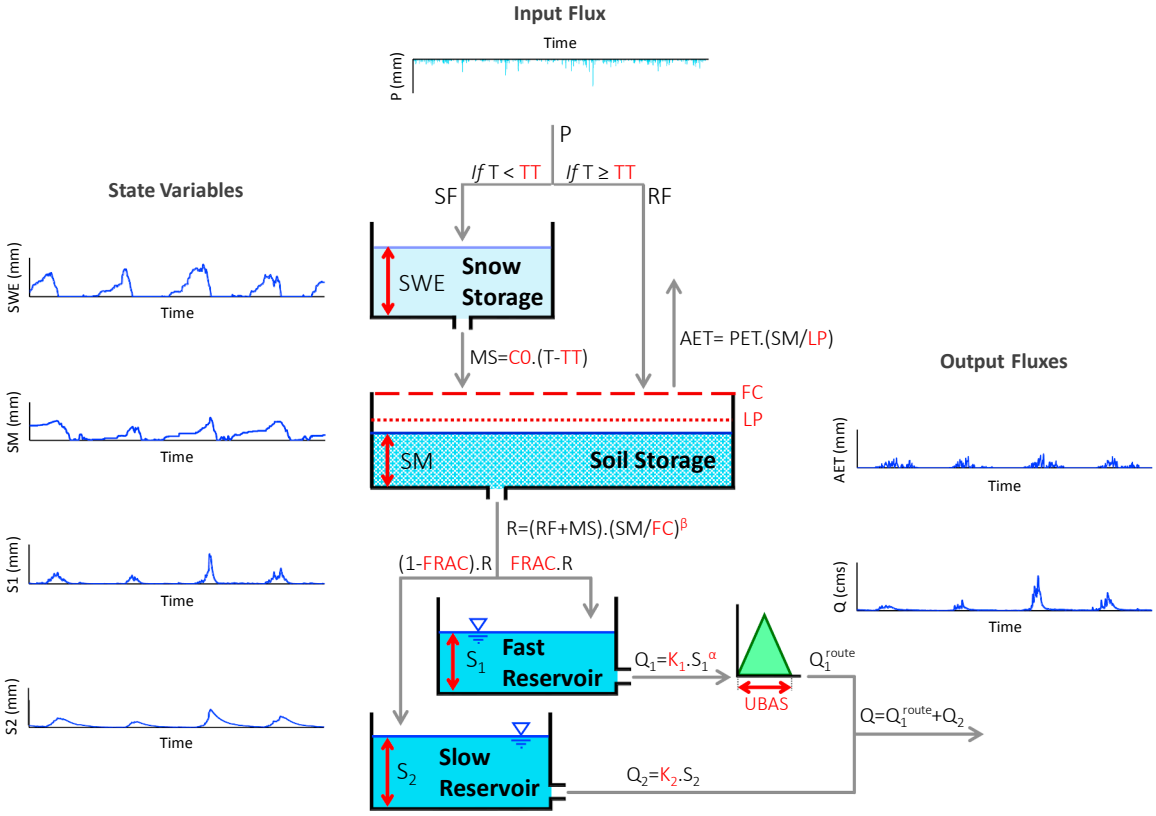

Fig. 9. The architecture of the HBV-SASK hydrologic model. The model works on a daily time step and reports all the flux and state variables. P: precipitation; T: temperature; SF: snowfall; RF: rainfall; SWE: snow water equivalent; MS: melted snow; AET: actual evapotranspiration; PET: potential evapotranspiration; SM: soil moisture; R: soil release; S_1 and S_2: storage in fast and slow reservoirs; Q_1 and Q_2: discharge from the fast and slow reservoirs; Q_1^{route}: discharge Q_1 routed by the watershed unit hydrograph; Q: total watershed discharge.

Table 2
The parameters of the HBV-SASK hydrologic model.

Number	Parameter Name	Lower Bound	Upper Bound	Description
1	TT	−4	4	Air temperature threshold in °C for melting/freezing and separating rain and snow
2	C0	0	10	Base melt factor, in mm/°C per day
3	ETF	0	1	Temperature anomaly correction in 1/°C of potential evapotranspiration
4	LP	0	1	Limit for PET as a multiplier to FC, i.e., soil moisture below which evaporation becomes supply limited
5	FC	50	500	Field capacity of soil, in mm. The maximum amount of water that the soil can retain
6	β (beta)	1	3	Shape parameter (exponent) for soil release equation (unitless)
7	FRAC	0.1	0.9	Fraction of soil release entering fast reservoir (unitless)
8	K1	0.05	1	Fast reservoir coefficient, which determines what proportion of the storage is released per day (unitless)
9	α (alpha)	1	3	Shape parameter (exponent) for fast reservoir equation (unitless)
10	K2	0	0.05	Slow reservoir coefficient which determines what proportion of the storage is released per day (unitless)
11	UBAS	1	3	Base of unit hydrograph for watershed routing in day; default is 1 for small watersheds
12	PM	0.5	2	Precipitation multiplier to address uncertainty in precipitation (unitless); default is 1.

* (T – T_{mth})).* PET_{mth}, where T_{mth} and PET_{mth} are long-term average monthly temperature and potential evapotranspiration, respectively, supplied as input data to the model along with daily precipitation (P) and temperature (T). Table 2 lists the model parameters (including ETF in the equation above), their description, and feasible ranges. The implementation is modular, with each module simulating a different hydrologic process, thereby allowing the user to easily investigate the different model flux and state variables (see Fig. 10).

HBV-SASK in VARS-TOOL comes with "ready-to-run" case studies for two watersheds, Bow River (at Banff with area of 2178.53 km^2) and Oldman River (at Waldron's Corner with an area of 1434.73 km^2). These rivers are located in the Rocky Mountains in Alberta, Canada, and flow into the Saskatchewan River Basin (Fig. 11). Historical data is available for the periods 1950–2011 and 1979–2008 respectively, from which we estimate average annual precipitation (rainfall + snowfall) to be 795 mm (Bow) and 611 mm (Oldman), and average annual streamflow to be 38.6 m^3/s at gauge 05BB001 on the Bow River, and 11.7 m^3/s at gauge 05AA023 on the Oldman River. The Bow and Oldman basins differ in their hydrological properties and have runoff ratios of approximately 0.7 and 0.42, respectively.

VARS-TOOL also comes with a ready-to-use set-up of the MESH (Modélisation Environmentale communautaire – Surface & Hydrology) land surface-hydrology model (Pietroniro et al., 2006). MESH couples the Canadian Land Surface Scheme (CLASS) (Verseghy et al., 1993; Verseghy, 1991) with the hydrologic routing schemes of WATFLOOD (Kouwen et al., 1993). This model set-up is interfaced with VARS-TOOL via the OSTRICH toolkit (Matott, 2017). The case studies with MESH are adopted from Haghnegahdar et al. (2017). Details on how to run this case study are available in the VARS-TOOL manual.

10. Concluding remarks

Significant theoretical and practical advances have been made in recent years in the field of sensitivity and uncertainty analyses. These analyses need to become an integral part of any model development, prediction, and decision-making process, providing insight into various issues such as uncertainty apportionment, diagnostic testing, planning and management, and policy prioritization. Best practices are, however, often hampered by computational burden and lack of transparency and interpretability. VARS-TOOL is designed as a computationally-efficient

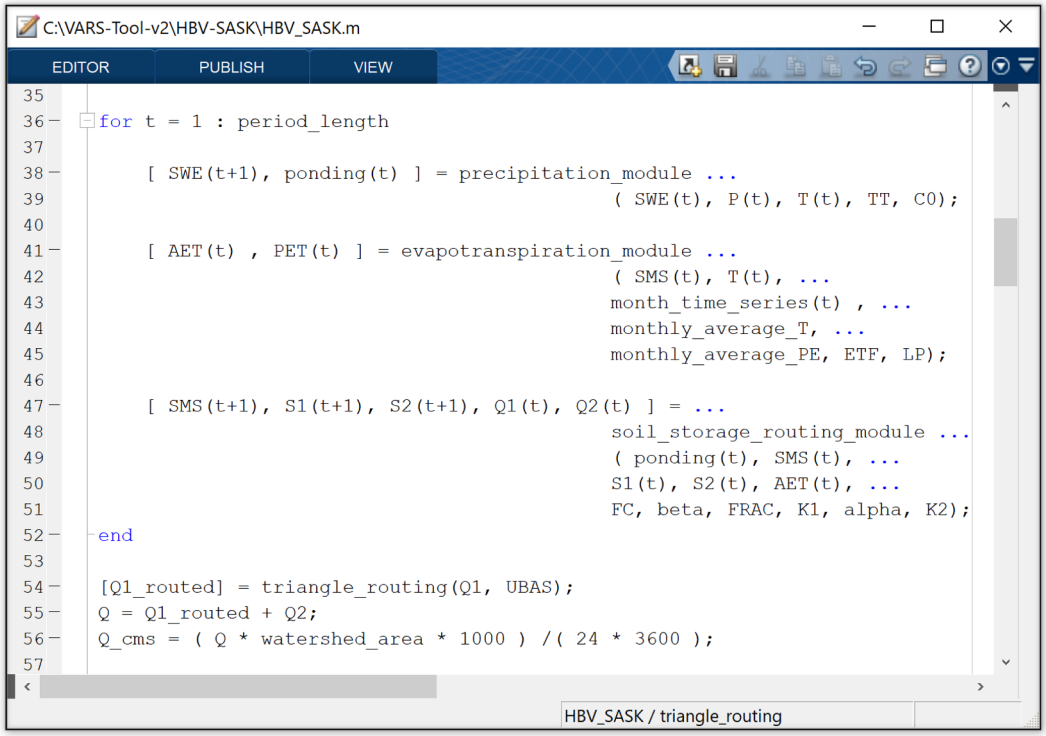

Fig. 10. A screenshot of the modules of the HBV-SASK rainfall-runoff model.

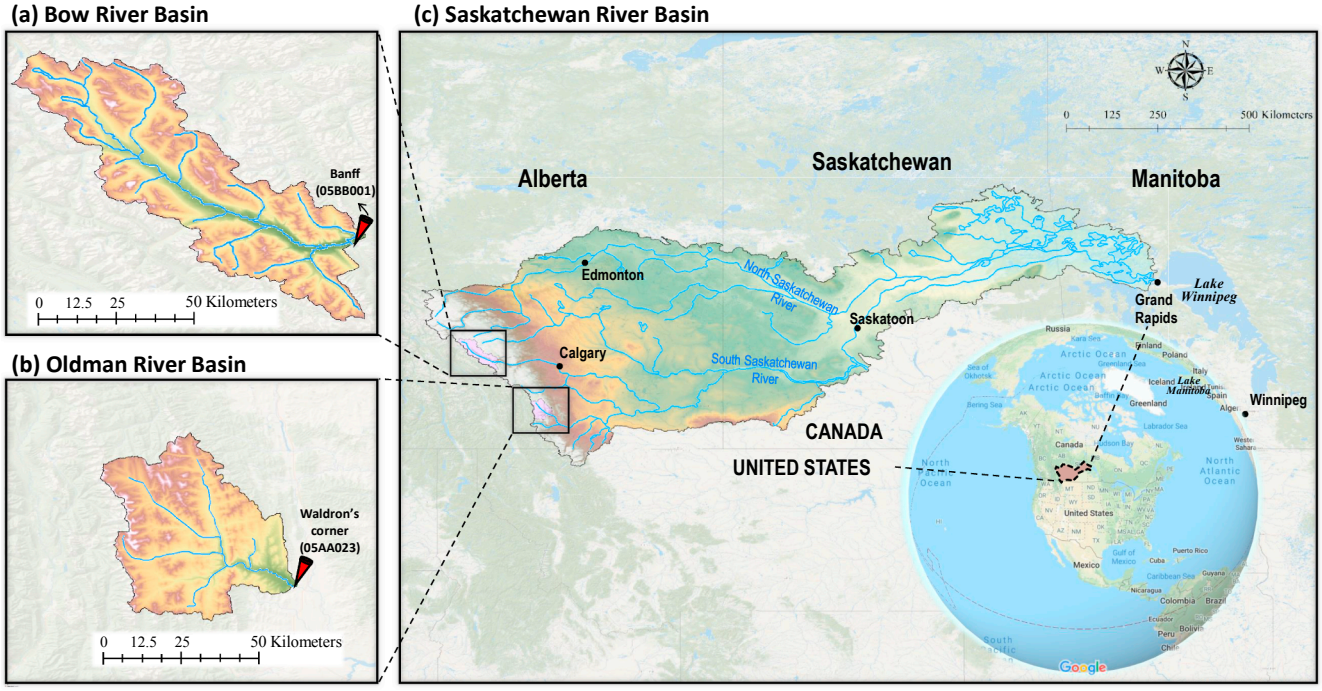

Fig. 11. Maps of the (a) Bow and (b) Oldman River Basins and their locations in (c) the Saskatchewan River Basin, Canada.

and transparent platform to reconcile the state-of-the-art theories with real-world applications that involve complex, high-dimensional, and computationally expensive models of dynamical systems.

In summary, VARS-TOOL aims to promote best practices in sensitivity analysis that should be based on:

- A multi-method approach that brings together the different theories and aspects of sensitivity analysis, thereby providing a more comprehensive assessment of the problem at hand; using, for example, the Variogram Analysis of Response Surfaces (VARS) approach (see Section 3).
- Proper accounting for the dynamical properties of Earth and environmental systems models; using, for example, the Generalized Global Sensitivity Matrix (GGSM) approach (see Section 4).
- Proper sampling of the problem space that effectively scales with the size of the problem at hand and available computational budget; using, for example, progressive Latin hypercube sampling (PLHS, see Section 5).
- Methods that are capable of handling high-dimensional problems, such as advanced Earth and environmental systems models that tend to involve hundreds of factors whose influence on model response needs to be characterized; using, for example, a factor-grouping strategy (see Section 6).
- Proper characterization of "robustness" (and stability) of the algorithms to sampling variability, and of the degree of "confidence" and "reliability" that users can have in sensitivity results; Using, for example, a bootstrap technique (see Section 7).

VARS-TOOL is under continuous development to include new theoretical advances in the field of sensitivity and uncertainty analysis. Forthcoming additions will include (but are not limited to):

(a) The ability to carry out sensitivity analysis on any sample data of a process, however collected (see e.g., Borgonovo et al., 2016). We believe this ability is needed for two reasons: (i) the utility of sensitivity analysis may not be limited to models and their artifacts, and one may directly apply sensitivity analysis to data to characterize the behavior of the underlying system processes; (ii) most sensitivity analysis algorithms require samples taken in a specific manner, which limits their utility when a sample set already exists.

(b) The ability to handle problems with non-uniform and/or correlated factors (see e.g., Kucherenko et al., 2012). Most algorithms for sensitivity analysis assume uniform a priori distributions for the factors. This assumption limits the applicability and validity of results, if the true distribution of factors is non-uniform and/or correlated.

(c) The ability to provide a more informative characterization of "interaction" effects (see e.g., Razavi and Gupta, 2015). Very few existing approaches to sensitivity analysis can provide a meaningful and/or easily interpretable assessment of interaction effects. There is a need to rethink the fundamentals of interaction effects, in relationship with the correlation effects (point b above) and to develop more informative tools for their characterization in support of model development and prediction.

The field of sensitivity analysis is still young and evolving. It has the potential to provide a solid foundation based on which different sources of uncertainty, which may by essence be irreconcilable, can be compared and assessed. Further, it has the potential to help us isolate, assess, and deal with the "deep uncertainties" in our future Earth and environmental systems that occur in the context of climatic, environmental, and social change. As always, we invite discussion and collaboration on any of these and other issues related to diagnostic evaluation and improvement of dynamical systems models, especially with regard to high-dimensional representations of complex systems.

Acknowledgements

The first, second, and fourth authors were supported in part by the first author's NSERC (Natural Sciences and Engineering Research Council of Canada) Discovery Grant. The third author received partial support from the Australian Research Council through the Centre of Excellence for Climate System Science (grant CE110001028).

References

Adams, B.M., Bohnhoff, W., Dalbey, K., Eddy, J., Eldred, M., Gay, D., Haskell, K., Hough, P.D., Swiler, L.P., 2009. DAKOTA, a Multilevel Parallel Object-Oriented Framework for Design Optimization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: version 5.0 User's Manual. Sandia National Laboratories Tech. Rep. SAND2010-2183.

Bennett, N.D., Croke, B.F., Guariso, G., Guillaume, J.H., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, S., Newham, L.T., Norton, J.P., Perrin, C., 2013. Characterising performance of environmental models. Environ. Model. Software 40, 1–20.

Beran, R., 1997. Diagnosing bootstrap success. Ann. Inst. Stat. Math. 49 (1), 1–24.

Beven, K., Binley, A., 1992. The future of distributed models: model calibration and uncertainty prediction. Hydrol. Process. 6 (3), 279–298.

Borgonovo, E., Hazen, G.B., Plischke, E., 2016. A common rationale for global sensitivity measures and their estimation. Risk Anal. 36 (10), 1871–1895.

Box, G.E., Meyer, R.D., 1986. An analysis for unreplicated fractional factorials. Technometrics 28 (1), 11–18.

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

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

Davison, A.C., Hinkley, D.V., Young, G.A., 2003. Recent developments in bootstrap methodology. Stat. Sci. 141–157.

Dell'Oca, A., Riva, M., Guadagnini, A., 2017. Moment-based metrics for global sensitivity analysis of hydrological systems. Hydrol. Earth Syst. Sci. 21 (12), 6219.

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

Efron, B., 1992. Bootstrap Methods: Another Look at the Jackknife, Breakthroughs in Statistics. Springer, pp. 569–593.

Gan, Y., Duan, Q., Gong, W., Tong, C., Sun, Y., Chu, W., Ye, A., Miao, C., Di, Z., 2014. A comprehensive evaluation of various sensitivity analysis methods: a case study with a hydrological model. Environ. Model. Software 51, 269–285.

Gupta, H.V., Razavi, S., 2018. Revisiting the basis of sensitivity analysis for dynamical earth system models. Water Resour. Res. 54. https://doi.org/10.1029/2018WR022668.

Haghnegahdar, A., Razavi, S., 2017. Insights into sensitivity analysis of earth and environmental systems models: on the impact of parameter perturbation scale. Environ. Model. Software 95, 115–131.

Haghnegahdar, A., Razavi, S., Yassin, F., Wheater, H., 2017. Multicriteria sensitivity analysis as a diagnostic tool for understanding model behaviour and characterizing model uncertainty. Hydrol. Process. 31 (25), 4462–4476.

Halton, J.H., 1960. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. Numer. Math. 2 (1), 84–90.

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

Hong, Y., Hsu, K.l., Moradkhani, H., Sorooshian, S., 2006. Uncertainty quantification of satellite precipitation estimation and Monte Carlo assessment of the error propagation into hydrologic response. Water Resour. Res. 42 (8).

Ishigami, T., Homma, T., 1990. An importance quantification technique in uncertainty analysis for computer models, Uncertainty Modeling and Analysis, 1990. In: Proceedings., First International Symposium on. IEEE, pp. 398–403.

Jamil, M., Yang, X.-S., 2013. A literature survey of benchmark functions for global optimisation problems. Int. J. Math. Model. Numer. Optim. 4 (2), 150–194.

JRC, 2008. SimLab 2.2 REFERENCE MANUAL. https://ec.europa.eu/jrc/en/samo/simlab.

Kaizer, J.S., Heller, A.K., Oberkampf, W.L., 2015. Scientific computer simulation review. Reliab. Eng. Syst. Saf. 138, 210–218.

Kouwen, N., Soulis, E., Pietroniro, A., Donald, J., Harrington, R., 1993. Grouped response units for distributed hydrologic modeling. J. Water Resour. Plann. Manag. 119 (3), 289–305.

Kucherenko, S., Tarantola, S., Annoni, P., 2012. Estimation of global sensitivity indices for models with dependent variables. Comput. Phys. Commun. 183 (4), 937–946.

Kunstmann, H., Kinzelbach, W., Siegfried, T., 2002. Conditional first-order second-moment method and its application to the quantification of uncertainty in groundwater modeling. Water Resour. Res. 38 (4).

Kwakkel, J.H., Haasnoot, M., Walker, W.E., 2016. Comparing robust decision-making and dynamic adaptive policy pathways for model-based decision support under deep uncertainty. Environ. Model. Software 86, 168–183.

Laguna, M., Martí, R., 2005. Experimental testing of advanced scatter search designs for global optimization of multimodal functions. J. Global Optim. 33 (2), 235–255.

Lindström, G., Johansson, B., Persson, M., Gardelin, M., Bergström, S., 1997. Development and test of the distributed HBV-96 hydrological model. J. Hydrol. 201 (1–4), 272–288.

Maier, H.R., Guillaume, J.H., van Delden, H., Riddell, G.A., Haasnoot, M., Kwakkel, J.H., 2016. An uncertain future, deep uncertainty, scenarios, robustness and adaptation: how do they fit together? Environ. Model. Software 81, 154–164.

Marelli, S., Sudret, B., 2014. UQLab: a framework for uncertainty quantification in matlab, vulnerability, uncertainty, and risk: quantification. Mitigation, and Management 2554–2563.

Matott, L., 2017. OSTRICH: an Optimization Software Tool, Documentation and User's Guide, Version 17.12.19. University at Buffalo Center for Computational Research, pp. 79. www.eng.buffalo.edu/~lsmatott/Ostrich/OstrichMain.html.

McKay, M.D., Beckman, R.J., Conover, W.J., 1979. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21 (2), 239–245.

Morris, M.D., 1991. Factorial sampling plans for preliminary computational experiments. Technometrics 33 (2), 161–174.

Muleta, M.K., Nicklow, J.W., 2005. Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model. J. Hydrol. 306 (1–4), 127–145.

Oreskes, N., 2003. In: Canham, C.D., Cole, J.J., Lauenroth, W.K. (Eds.), The Role of Quantitative Models in Science Naomi Oreskes. Models in Ecosystem Science, pp. 13–31.

Pianosi, F., Sarrazin, F., Wagener, T., 2015. A Matlab toolbox for global sensitivity analysis. Environ. Model. Software 70, 80–85.

Pianosi, F., Wagener, T., 2015. A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. Environ. Model. Software 67, 1–11.

Pietroniro, A., Fortin, V., Kouwen, N., Neal, C., Turcotte, R., Davison, B., Verseghy, D., Soulis, E., Caldwell, R., Evora, N., 2006. Using the MESH modelling system for hydrological ensemble forecasting of the Laurentian Great Lakes at the regional scale. Hydrol. Earth Syst. Sci. Discuss. 3 (4), 2473–2521.

Rakovec, O., Hill, M.C., Clark, M., Weerts, A., Teuling, A., Uijlenhoet, R., 2014. Distributed evaluation of local sensitivity analysis (DELSA), with application to hydrologic models. Water Resour. Res. 50 (1), 409–426.

Razavi, S., 2017. Feature: when Uncertainty Matters. American Geophysical Union-Hydrology Section December 2017 Newsletter Invited Commentary. https://hydrology.agu.org/wp-content/uploads/sites/19/2017/11/HS-December-2017-Newsletter_final.pdf.

Razavi, S., Gupta, H.V., 2015. What do we mean by sensitivity analysis? The need for comprehensive characterization of "global" sensitivity in Earth and Environmental systems models. Water Resour. Res. 51 (5), 3070–3092.

Razavi, S., Gupta, H.V., 2016a. A new framework for comprehensive, robust, and efficient global sensitivity analysis: 1. Theory. Water Resour. Res. 52 (1), 423–439.

Razavi, S., Gupta, H.V., 2016b. A new framework for comprehensive, robust, and efficient global sensitivity analysis: 2. Application. Water Resour. Res. 52 (1), 440–455.

Razavi, S., Gupta, H.V., In Review. A General Approach to Multi-method Sensitivity Analysis of Dynamical Systems Models. Environmental Modelling & Software.

Razavi, S., Tolson, B.A., Burn, D.H., 2012. Review of surrogate modeling in water resources. Water Resour. Res. 48 (7).

Saltelli, A., Annoni, P., 2010. How to avoid a perfunctory sensitivity analysis. Environ. Model. Software 25 (12), 1508–1517.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2008. Global Sensitivity Analysis: the Primer. John Wiley & Sons.

Sheikholeslami, R., Razavi, S., 2017. Progressive Latin Hypercube Sampling: an efficient approach for robust sampling-based analysis of environmental models. Environ. Model. Software 93, 109–126.

Sheikholeslami, R., Razavi, S., Gupta, H.V., Becker, W., Haghnegahdar, A., 2018. Global sensitivity analysis for high-dimensional problems: how to objectively group factors and measure robustness and convergence while reducing computational cost. Environ. Model. Software. https://doi.org/10.1016/j.envsoft.2018.09.002.

Sheikholeslami, R., Razavi, S., Haghnegahdar, A., (In Prep). What Should We Do with Model Simulation Crashes? Recommendations for Global Sensitivity Analysis of Earth Systems Models.

Smith, T.J., Marshall, L.A., 2008. Bayesian methods in hydrologic modeling: a study of recent advancements in Markov chain Monte Carlo techniques. Water Resour. Res. 44 (12).

Sobol, I.y.M., 1967. On the distribution of points in a cube and the approximate evaluation of integrals. Zh. Vychislitel noi Mat. Mat. Fiz. 7 (4), 784–802.

Sobol, I., Levitan, Y.L., 1999. On the use of variance reducing multipliers in Monte Carlo computations of a global sensitivity index. Comput. Phys. Commun. 117 (1), 52–61.

Sobol, I.M., 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Math. Comput. Simulat. 55 (1–3), 271–280.

Sobol, I.M., Kucherenko, S., 2009. Derivative based global sensitivity measures and their link with global sensitivity indices. Math. Comput. Simulat. 79 (10), 3009–3017.

Spear, R.C., Grieb, T.M., Shang, N., 1994. Parameter uncertainty and interaction in complex environmental models. Water Resour. Res. 30 (11), 3159–3169.

Tetzlaff, D., McDonnell, J., Uhlenbrook, S., McGuire, K., Bogaart, P., Naef, F., Baird, A., Dunn, S., Soulsby, C., 2008. Conceptualizing catchment processes: simply too complex? Hydrol. Process. 22 (11), 1727–1730.

Touzani, S., Busby, D., 2014. Screening method using the derivative-based global sensitivity indices with application to reservoir simulator. Oil & Gas Science and Technology–Revue d'IFP Energies nouvelles 69 (4), 619–632.

Verseghy, D., McFarlane, N., Lazare, M., 1993. CLASS—a Canadian land surface scheme for GCMs, II. Vegetation model and coupled runs. Int. J. Climatol. 13 (4), 347–370.

Verseghy, D.L., 1991. CLASS—a Canadian land surface scheme for GCMs. I. Soil model. Int. J. Climatol. 11 (2), 111–133.

Vrugt, J.A., Gupta, H.V., Bouten, W., Sorooshian, S., 2003. A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. Water Resour. Res. 39 (8).

Welch, W.J., Buck, R.J., Sacks, J., Wynn, H.P., Mitchell, T.J., Morris, M.D., 1992. Screening, predicting, and computer experiments. Technometrics 34 (1), 15–25.

Yassin, F., Razavi, S., Wheater, H., Sapriza-Azuri, G., Davison, B., Pietroniro, A., 2017. Enhanced identification of a hydrologic model using streamflow and satellite water storage data: A multicriteria sensitivity analysis and optimization approach. Hydrol. Process. 31 (19), 3320–3333. https://doi.org/10.1002/hyp.11267.

Ye, K.Q., Li, W., Sudjianto, A., 2000. Algorithmic construction of optimal symmetric Latin hypercube designs. J. Stat. Plann. Inference 90 (1), 145–159.