AGU100 ADVANCING EARTH AND SPACE SCIENCE

Water Resources Research

RESEARCH ARTICLE
10.1029/2018WR022668

Key Points:
- Questions the use of the performance-metric-based sensitivity analysis of Dynamical Earth Systems Models and shows that the analysis it provides is both inaccurate and incomplete
- Theoretically frames the global sensitivity analysis problem from first principles and develops a performance metric-free approach to assessing parameter importance
- Demonstrates that the new approach is efficient, stable, and robust and disagrees with metric-based methods regarding which parameters exert the strongest controls on model behavior

01
Correspondence to:
H. V. Gupta,
hoshin.gupta@hwr.arizona.edu
mingxi zhang

Citation:
Gupta, H. V., & Razavi, S. (2018). Revisiting the basis of sensitivity analysis for Dynamical Earth System Models. *Water Resources Research*, *54*, 8692–8717. https://doi.org/10.1029/2018WR022668

Received 29 JAN 2018
Accepted 2 SEP 2018
Published online 8 NOV 2018

Revisiting the Basis of Sensitivity Analysis for Dynamical Earth System Models

Hoshin V. Gupta[1] **and Saman Razavi**[2,3]

[1]Department of Hydrology and Water Resources, University of Arizona, Tucson, Arizona, USA, [2]Global Institute for Water Security, University of Saskatchewan, Saskatoon, Saskatchewan, Canada, [3]School of Environment and Sustainability and Department of Civil and Geological Engineering, University of Saskatchewan, Saskatoon, Saskatchewan, Canada

Abstract This paper investigates the problem of global sensitivity analysis (GSA) of Dynamical Earth System Models and proposes a basis for how such analyses should be performed. We argue that (a) performance metric-based approaches to parameter GSA are actually identifiability analyses, (b) the use of a performance metric to assess sensitivity unavoidably distorts the information provided by the model about relative parameter importance, and (c) it is a serious conceptual flaw to interpret the results of such an analysis as being consistent and accurate indications of the sensitivity of the model response to parameter perturbations. Further, because such approaches depend on availability of system state/output observational data, the analysis they provide is necessarily incomplete. Here we frame the GSA problem from first principles, using trajectories of the partial derivatives of model outputs with respect to controlling factors as the theoretical basis for sensitivity, and construct a global sensitivity matrix from which statistical indices of whole-period time-aggregate *parameter importance*, and time series of time-varying parameter importance, can be inferred. We demonstrate this framework using the *HBV-SASK* conceptual hydrologic model applied to the Oldman basin in Canada and show that it disagrees with performance metric-based methods regarding which parameters exert the strongest controls on model behavior. Further, it is highly efficient, requiring less than 1,000 base samples to obtain stable and robust parameter importance assessments for our 10-parameter example.

Plain Language Summary When developing and using computer-based models to (a) understand Earth and environmental systems, (b) make predictions, and/or (c) make management or policy decisions, it is very important to know which factors most strongly control the behaviors of the model. Tools to determine this are called sensitivity analysis (SA) methods. This paper shows that the use of model performance metrics to assess sensitivity is based in faulty reasoning. By framing the problem from *first principles*, a logical approach is developed that provides accurate and cost-effective assessments of both time-aggregate and time-varying parameter importance. Because the approach does not require availability of system output data, it enables a comprehensive assessment and can be applied to historical and predictive conditions, as well as to future scenarios.

1. Introduction, Background, and Motivation

1.1. The Basis for Sensitivity Analysis of Dynamical Earth System Models

2-3
2 notes:

Dynamical Earth System Models (DESMs) that summarize and reflect our growing understanding about the world are rapidly becoming more complex and computationally intensive. As they do so, the need for robust, informative, and computationally efficient sensitivity analysis (SA) techniques and tools is becoming ever more pressing. Broadly, the most common purpose of SA is to establish which parameters in a model most strongly affect the magnitude, variability, and dynamics of model response (Razavi & Gupta, 2015). This enables an assessment of how much care must be taken in specifying the values of various model parameters when using the model in a decision making context. Therefore, parameter SA is most useful when performed independently of the model calibration process (typically before model calibration is carried out), and one typically requires a *global* sensitivity analysis (GSA) that is indicative of relative parameter importance regardless of what the *optimal* parameter set(s) may be. Our goal is to revisit the basis for how SA can be applied in the context of DESMs and, thereby, to contribute to the discussion of how SA can aid the model evaluation process.

04
mingxi zhang

To begin, we recognize that the basis for any SA is an assessment of how some model response $\boldsymbol{R}$ changes when some controlling factor $\boldsymbol{C}$ is altered. As discussed in section 1.2, $\boldsymbol{R}$ can be any quantity of interest, including (but not limited to) the model output, any statistical property thereof (e.g., the variance), or any characteristic property thereof (e.g., some signature property such as the long-term temporal trend, slope of the midpoint of the flow duration curve, or spatial trend). Similarly, $\boldsymbol{C}$ can be any factor that influences the properties of $\boldsymbol{R}$, including (but not limited to) the parameters, the parameterization equations, the boundary conditions of the system, the input fluxes, and/or the distributions of material properties, etc. While our presentation will focus mainly on the sensitivity of model states and output fluxes to perturbations of the parameters, SA can be applied to any combination of model response and controlling factor of interest.

Now, when $\boldsymbol{C}$ is perturbed by amount $\Delta\boldsymbol{C}$, such that the change in response is $\Delta\boldsymbol{R}$, it is usual to express the *per unit* sensitivity of $\boldsymbol{R}$ to $\boldsymbol{C}$ via a *sensitivity coefficient* $\boldsymbol{S} = \Delta\boldsymbol{R}/\Delta\boldsymbol{C}$ (or by its absolute value if only the magnitude, and not the sign, is considered to be important). Other expressions are also sometimes used, such as the normalized $\frac{\Delta R}{R} / \frac{\Delta C}{C}$. More generally, the change in $\boldsymbol{C}$ and/or $\boldsymbol{R}$ can be either *quantitative or qualitative*; for purposes of communication we will refer here to the changes as $\Delta\boldsymbol{R}$ and $\Delta\boldsymbol{C}$.

There are three possible cases:

Case 1) The perturbation $\Delta\boldsymbol{C}$ is selected (by the user) to be some finite amount, in which case the sensitivity coefficient is computed as $\boldsymbol{S} = \Delta\boldsymbol{R}/\Delta\boldsymbol{C}$ (or $|\Delta\boldsymbol{R}/\Delta\boldsymbol{C}|$) as indicated above.

Case 2) If we allow the perturbation $\Delta\boldsymbol{C} \rightarrow \boldsymbol{0}$, then $\boldsymbol{S}$ is expressed by the mathematical derivative $d\boldsymbol{R}/d\boldsymbol{C}$ (or $|d\boldsymbol{R}/d\boldsymbol{C}|$).

Case 3) If $\boldsymbol{C}$ can take on only several discrete conditions $\boldsymbol{\{Condition\ 1, Condition\ 2, \ldots, Condition\ N_c\}}$, we express $\boldsymbol{S}$ as the change $\Delta\boldsymbol{R}$ when $\boldsymbol{C}$ is varied from one condition to another.

In general, the value of $\boldsymbol{S}$ will vary throughout the factor space and should be indexed by location, so that $\boldsymbol{S_o}$ corresponds to a factor value/condition $\boldsymbol{C_o}$. A simple example of Case 3 is when evaluating several alternative discrete *land cover* types.

As best as we can tell, all existing approaches to SA fall within this classification. Case 2 represents the traditional mathematical interpretation of sensitivity and forms the theoretical basis for *derivative-based* approaches such as *Morris* (Campolongo et al., 2007; Morris, 1991) and *DELSA* (Rakovec et al., 2014). However, because computation of actual derivatives is often inconvenient, the more common approach is to use finite difference approximations and therefore corresponds to Case 1. Meanwhile, Case 3 is the basis for *variance-based* methods such as *Sobol'* (Saltelli et al., 2008; Sobol', 1990) and *FAST* (Cukier et al., 1978), in which $\boldsymbol{R}$ is the variance of some quantity generated by the system when $\boldsymbol{C}$ is varied randomly over its domain. In this *analysis of variance* approach, $\boldsymbol{S}$ is computed as the (normalized) change in variance when the condition $\boldsymbol{C}$ is changed from **Vary** to **Fixed** (i.e., $\boldsymbol{S} = (\boldsymbol{R_{C=Vary}} - \boldsymbol{R_{C=Fixed}})/\boldsymbol{R_{C=Vary}}$); here $\boldsymbol{C}$ = **Vary** corresponds to allowing $\boldsymbol{C}$ to vary over its domain of interest and $\boldsymbol{C}$ = **Fixed** corresponds to fixing the factor to some specific value $\boldsymbol{C_i}$ within that domain. Because $\boldsymbol{C_i}$ can be anywhere within the domain of $\boldsymbol{C}$, the quantity $\boldsymbol{R_{C=Fixed}}$ is typically computed as the *average* of the values $\boldsymbol{R_i}$ obtained when $\boldsymbol{C_i}$ is (randomly) fixed at different possible values across its domain.

Two comments are worth making here. First, whereas *Sobol'* and *FAST* rely on changes in the *variance* of $\boldsymbol{R}$ as the sensitivity measure of interest, other statistical quantities can also be used, including changes in higher-order moments (skewness and kurtosis; Dell'Oca et al., 2017) or changes in properties of the cumulative distribution function (Pianosi & Wagener, 2015). Second, both derivative-based (e.g., *Morris*) and variance-based (e.g., *Sobol'*) approaches exist as special cases of the Variogram Analysis of Response Surfaces (VARS; Razavi & Gupta, 2016a, 2016b) approach when the perturbation size $\Delta\boldsymbol{C}$ is varied from infinitesimally small ($\Delta\boldsymbol{C} \rightarrow \boldsymbol{0}$) to very large ($\Delta\boldsymbol{C} \rightarrow \boldsymbol{Range_C}$). In other words, the response surface variogram provides a spectrum of sensitivity relevant information, ranging from *Morris* to *Sobol'* and other scale interpretations in between.

1.2. The Choice of Response Used in the Evaluation of Sensitivity

Typically, GSA approaches are designed to provide information about the sensitivity of a *single* user-selected model response $\boldsymbol{R}$ (Morris, 1991; Sobol' & Levitan, 1999; Sobol' & Kucherenko, 2009). DESMs, however, generate spatiotemporal dynamical responses in the form of state variables $\boldsymbol{X}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},\boldsymbol{t})$ and fluxes $\boldsymbol{Y}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},\boldsymbol{t})$ that evolve through time ($\boldsymbol{t}$), and that can exist over three-dimensional domains ($\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}$) ranging from lumped

(aggregate average over the domain) to fully three-dimensional. Our review of the literature indicates that GSA studies typically define the model *response* ***R*** in one of the following four ways (see also Shin et al., 2013, and Pianosi et al., 2016, for discussion of the first two), using the following:

Method 1) Model performance metrics that quantify the closeness of the dynamic state-flux response to observed data
Method 2) A specific targeted aspect of the dynamic state-flux response
Method 3) A compressed set of properties that characterize the dynamic state-flux response
Method 4) Direct reference to the spatiotemporally varying dynamic state-flux responses

1.2.1. Use of Model Performance Metrics

In hydrological modeling, perhaps the most frequent reason for conducting SA is to select the *most sensitive* parameters to vary (alternatively the *least sensitive* parameters to fix) during model calibration. In such studies, it is common for ***R*** to be defined in terms of some summary metric of model performance that quantifies goodness of fit of the model response to observed data (e.g., Borgonovo et al., 2017; Cloke et al., 2008; Demaria et al., 2007; Haghnegahdar & Razavi, 2017; Pappenberger et al., 2008; Rosolem et al., 2012; van Griensven et al., 2006; van Werkhoven et al., 2008b). Recognizing this approach to be poorly informative given the dynamical nature of the system, extensions to account for system dynamics have been proposed (e.g., see Cibin et al., 2010; Herman, Reed, et al., 2013; Herman, Kollat, et al., 2013; Massmann et al., 2014; Pianosi & Wagener, 2016; Sieber & Uhlenbrook, 2005; van Werkhoven et al., 2008a). In general, these extensions apply SA to the model performance metric computed on sequential time windows of the model response (a moving window approach). As discussed below (sections 1.3 and 1.4), we do not recommend this approach if the goal of SA is to support better understanding of the system/model.

1.2.2. Use of a Single Targeted Aspect of Model Response

Instead of being concerned with the entire spatiotemporal nature of the dynamical system response, attention can be simply devoted to some specific aspect of that response. For example, van Griensven et al. (2006) used the average catchment outlet streamflow over a period of time, Savage et al. (2016) used the spatiotemporally averaged maximum water depth, and Shin et al. (2013) used the number of days during which streamflow remained below a certain threshold. The value of this approach lies in its focus on quantities of relevance to some specific decision context. Unlike Method 1, the approach does not require the availability of observed data.

1.2.3. Use of a Compressed Set of Properties That Characterize the Dynamic State-Flux Response

Extending Method 2, a parallel line of research seeks to compress the spatiotemporal state-flux response of the system into a set of properties that characterizes that response. Unlike Method 2, this approach attempts to preserve (as well as possible) the complete information content of the system response (or at least the most important aspects thereof). Treating these characteristic properties as the responses ***R*** of interest, SA is then applied to assess how they vary with perturbations of the factors ***C***. For example, the seminal paper by Campbell et al. (2006) used a *data-driven* approach to investigate questions such as "What shifts the curves up and down or moves them left or right?" and "What makes the central peak wider or narrower?"; they transformed the model output into a new coordinate system and used GSA to assess the sensitivity of the most significant coefficients to perturbations of the model parameters. Such transformations can be carried out in numerous ways, including principal component analysis (Lamboni et al., 2009, 2011) and wavelet analysis (Marrel et al., 2011). In the hydrological/environmental sciences, a *physics-driven* approach has gained traction, which instead seeks to employ diagnostic signature properties of system response (Gupta et al., 2008) for model identification (although perhaps not yet in the context of SA). The potential advantage of this approach is that such signatures can be selected to have intuitive physical meaning (e.g., Moench, 1994; Rupp & Selker, 2006; Sivapalan et al., 2003; Vogel & Sankarasubramanian, 2003; Yilmaz et al., 2008, and many more) thereby providing support for hypothesis testing.

1.2.4. Direct Use of Spatiotemporally Varying Dynamic State-Flux Responses

The most detailed approach to SA selects ***R*** to be the state-flux response at every time step or/and spatial location. Examples include study of the temporal dynamics of sensitivity of model generated streamflow (Guse et al., 2014; Reusser et al., 2011; Reusser & Zehe, 2011) to identify the dominant model components at different times and under different conditions and to assess the sensitivity of spatial flood maps to different sources of uncertainty (Abily et al., 2016). In an interesting study, Le Cozannet et al. (2015) investigated the time-varying dominance in sources of uncertainty for projections of the yearly probability of coastal

flooding over the next two hundred years. In keeping with our position that the most important goal of SA is to support better understanding of the system/model, the methods developed in this paper will focus on this fourth approach.

1.3. Implications of the Choices of Response and Analytical Method

Clearly, differences in choice of $\boldsymbol{R}$ (section 4.2) can lead to different assessments regarding which factors exert the strongest (weakest) controls on model behavior. Equally clearly, selection of what serves as $\boldsymbol{R}$ should depend on the objectives of the analysis. However, having selected $\boldsymbol{R}$, it is important to note that differences in the *method* used to assess sensitivity (section 1.1), for example *derivative*- versus *variance-based*, can also result in different conclusions regarding factor importance. The reason for this was discussed by Razavi and Gupta (2015) and Razavi and Gupta (2016a, 2016b) who demonstrated the importance of accounting for the covariance structure in model response as factors are varied (a fact that forms the basis for the VARS approach). If this is not done, it is possible (even likely) that different GSA methods can (undesirably) provide identical sensitivity rankings for situations having very different factor sensitivity properties.

So while *Sobol'* and *Morris* are (arguably) currently the methods most widely used for GSA of *DESMs*, we have previously argued that *VARS* represents the most comprehensive and powerful method discovered to date, because it (a) accounts for the *spatial* covariance in model response as factors are varied, (b) is relatively more efficient and cost effective than other methods (requiring smaller numbers of model runs), and (c) provides more reliable (stable) estimates in the face of sampling variability. In addition, due to its theoretical connection to *Morris* and *Sobol'*, *VARS* can generate reliable estimates of the *Sobol'* and *Morris'* sensitivity rankings as by-products at no extra computational expense.

Returning again to the choice of model response, it is important to be clear that when $\boldsymbol{R}$ is defined in terms of a goodness-of-fit model performance metric, what is actually being conducted is a form of factor identifiability analysis rather than a factor *sensitivity analysis*, in the sense that the analysis tells us more about which factors can be perturbed to improve model *fit* to the observed data (the calibration problem) rather than providing diagnostically useful information regarding model behavior that can be used to guide model improvement (the diagnostic problem). We explain this distinction further in the next section.

1.4. Sensitivity Versus Identifiability (The Filtering Role of Objective Functions)

Consider the case where $\boldsymbol{R}$ is defined to be some metric (sometimes more than one; see Rosolem et al., 2012) of aggregate model performance (let us call it $\boldsymbol{F}$). Examples of such metrics include the mean squared error (MSE), its related normalization the Nash-Sutcliffe Efficiency (Nash & Sutcliffe, 1970), and the recently proposed Kling-Gupta Efficiency (Gupta et al., 2009) and its mean bias, variability bias, and correlation components.

In this case, applications of *Sobol'* base their assessment of relative factor importance on a decomposition of factor contributions to the variance of the distribution $\boldsymbol{p}(\boldsymbol{F}^{\boldsymbol{j}})$ of performance metric values, while applications of *Morris'* use the means and standard deviations of the distributions $\boldsymbol{p}(\boldsymbol{dF}^{\boldsymbol{j}}/\boldsymbol{d\theta_i})$ of approximate partial derivatives of the performance metric with respect to the factors, where the distributions are obtained by randomly sampling a large number of factor locations (indexed by $\boldsymbol{j}$) uniformly across the feasible factor space. Similarly, *PAWN*, *DELSA*, *FAST,* and *VARS* are typically based on distributions of sampled metric values, their partial derivatives, or both (in case of VARS).

It is important, however, to distinguish between *sensitivity* and *identifiability*. For clarity, we use the term factor sensitivity analysis to refer to the process whereby we seek to establish which factors exert stronger (or weaker) controls on the models' dynamical input-state-output behavior; this is a specific attribute of the *forward* (simulation) problem. In contrast, factor identifiability analysis is an attribute of the *inverse* problem in which we seek to establish which factors are more readily identifiable (i.e., whose *best* values are more easily determined) when time series of observational data regarding the system behavior is available. These two attributes are not identical; while factor identifiability clearly depends on the sensitivity of the model's dynamical state-output response to factor perturbations and inputs, the converse is not true.

From our review of the literature, it appears that this distinction is not as well recognized and that what is commonly called a factor (usually parameter) sensitivity analysis is actually an identifiability analysis. As discussed below, it is a serious conceptual flaw to interpret the results of a performance metric-based

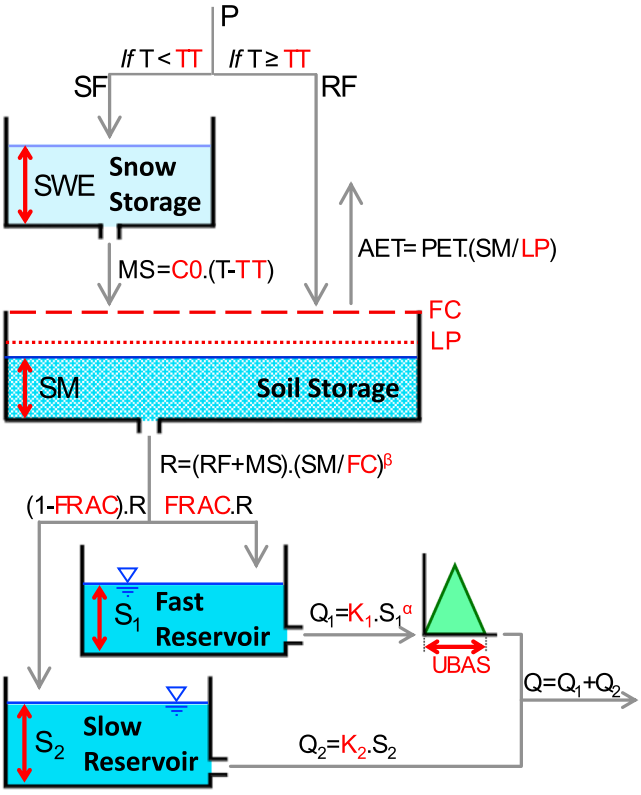

Figure 1. System architecture of the HBV-SASK hydrologic model.

identifiability analysis as being accurately indicative of the sensitivity of the model response to factor (parameter) perturbations.

1.4.1. Mathematical Formulation

Consider a DESM, driven by an input sequence $\boldsymbol{U_t}=\{\boldsymbol{U_t^1},\ldots,\boldsymbol{U_t^{D_u}}\}$, from an initial state $\boldsymbol{X_0}=\{\boldsymbol{X_0^1},\ldots,\boldsymbol{X_0^{D_x}}\}$, to give rise to a sequence of states $\boldsymbol{X_t}=\{\boldsymbol{X_t^1},\ldots,\boldsymbol{X_t^{D_x}}\}$ and outputs $\boldsymbol{Y_t}=\{\boldsymbol{Y_t^1},\ldots,\boldsymbol{Y_t^{D_y}}\}$, respectively, from time $\boldsymbol{t}=\boldsymbol{1}$ to $\boldsymbol{T}$. Here $\boldsymbol{D_u}$, $\boldsymbol{D_x}$, and $\boldsymbol{D_y}$ are the dimensions of the input, state, and output vectors, respectively. For example, the *HBV-SASK* hydrologic model (Figure 1) has two inputs (precipitation and long-term average monthly potential evapotranspiration), four state variables, and two outputs (streamflow and actual evapotranspiration).

At any time $\boldsymbol{t}$, the outputs $\boldsymbol{Y_t}$ depend on the input sequence $\boldsymbol{U_{1\rightarrow t}}=\{\boldsymbol{U_1},\ldots,\boldsymbol{U_t}\}$ from all prior time steps, the initial state $\boldsymbol{X_0}$, and the model *structure* and *parameters* indicated here by $\boldsymbol{M_\theta}(\cdot)$, where the model structure results from progressive specification of (Gupta & Nearing, 2014) the following:

1. The control volume (***CV***) and its boundary conditions (***BC***)
2. The system architecture (***SArch***, represented as a directed graph consisting of nodes and links, where the nodes correspond primarily to the state variables, sometimes called stocks, and the links correspond to the fluxes, sometimes called flows)
3. The process parameterization equations (***PPEs***, hereafter called $\boldsymbol{\psi}=\{\boldsymbol{\psi_1},\ldots,\boldsymbol{\psi_{N_\psi}}\}$) that make up the links, and
4. The process equation parameters (***PEPs***, hereafter called $\boldsymbol{\Theta}=\{\boldsymbol{\theta_1},\ldots,\boldsymbol{\theta_{N_\theta}}\}$), whose meanings depend on the specific selection of ***PPEs*** mentioned above; in other words, $\boldsymbol{\Theta}$ comes into being as a consequence of the selection of $\boldsymbol{\psi}$.

Here $\boldsymbol{N_\psi}$ and $\boldsymbol{N_\theta}$ are the numbers of ***PPEs*** and ***PEPs***, respectively (typically assumed to be constant across all time steps).

So the values of outputs $\boldsymbol{Y_t}$ depend on (are sensitive to variations in) $\boldsymbol{U_{1\rightarrow t}}$, $\boldsymbol{X_0}$, ***CV***, ***BC***, ***SArch***, $\boldsymbol{\psi}$, and $\boldsymbol{\Theta}$. In DES modeling, it is common to assume that ***CV***, ***BC***, ***SArch***, and $\boldsymbol{\psi}$ are fixed (i.e., known precisely) but that $\boldsymbol{U_{1\rightarrow t}}$, $\boldsymbol{X_0}$ and $\boldsymbol{\Theta}$ are known only imprecisely. Consequently, our interest is typically in sensitivity of $\boldsymbol{Y_t}$ to variations (imprecision in) in $\boldsymbol{U_{1\rightarrow t}}$, $\boldsymbol{X_0}$ and $\boldsymbol{\Theta}$.

Without loss of generality, the following discussion will focus on sensitivity of $\boldsymbol{Y_t}$ to variations in the different parameters $\boldsymbol{\theta_i}$; in other words, we are interested in characterizing how $\boldsymbol{Y_t}$ is affected by changes to the values of $\boldsymbol{\theta_i}$. In particular, we are interested in knowing which are the most *influential/important* parameters, in the sense that they exert the strongest controls on the behavior $\boldsymbol{Y_t}$, and, in general, we may wish to rank this importance from largest to smallest. Further, we are more generally concerned with the sensitivity of the *entire* temporal sequence of system outputs $\boldsymbol{Y}=\{\boldsymbol{Y_1},\ldots,\boldsymbol{Y_T}\}$ to variations in the parameters. Therefore, we represent this dependence as $\boldsymbol{Y}(\boldsymbol{\Theta})=\{\boldsymbol{Y_1}(\boldsymbol{\Theta}),\ldots,\boldsymbol{Y_T}(\boldsymbol{\Theta})\}$. Finally, we assume that the parameters are restricted to take on values within some feasible space $\boldsymbol{\Phi_\Theta}$.

1.4.2. On the Use of Model Performance Metrics for So-called Sensitivity Analysis

Consider, as an example, that we choose a typical MSE-type scalar metric $\boldsymbol{F^j}$ as our measure of model performance with regards to a specific output $\boldsymbol{Y}$, where the superscript $\boldsymbol{j}$ indicates that the metric is being evaluated at the parameter location $\boldsymbol{\Theta^j}$. Mathematically, we have

$$\boldsymbol{F^j}=\boldsymbol{F}\left(\boldsymbol{Y}|\boldsymbol{Z},\boldsymbol{\Theta^j}\right)=\frac{\boldsymbol{1}}{\boldsymbol{T}}\sum_{\boldsymbol{t=1}}^{\boldsymbol{T}}\left(\boldsymbol{fZ_t}-\boldsymbol{fY_t}\left(\boldsymbol{\Theta^j}\right)\right)^{\boldsymbol{2}} \quad (1)$$

where $\boldsymbol{Y}(\boldsymbol{\Theta^j})=\{\boldsymbol{Y_1}(\boldsymbol{\Theta^j}),\ldots,\boldsymbol{Y_T}(\boldsymbol{\Theta^j})\}$ is the time series of model outputs generated using parameter values $\boldsymbol{\Theta^j}$, $\boldsymbol{Z}=\{\boldsymbol{Z_1},\ldots,\boldsymbol{Z_T}\}$ is the corresponding time series of target (observed) outputs, and the operator $\boldsymbol{f}$ represents a transformation applied to the model generated and target output values before the

metric is computed; for example, with f as the natural log transform we have $F^j = \frac{1}{T}\sum_{t=1}^{T}\left(\ln Z_t - \ln Y_t(\Theta^j)\right)^2$.

Taking the local derivative of F^j with respect to each of the parameters $\Theta = \{\theta_1, \ldots, \theta_{N_\theta}\}$, we get the vector of partial derivatives:

$$\nabla F^j = \left\{dF^j/d\theta_1, \ldots, dF^j/d\theta_{N_\theta}\right\} \tag{2}$$

whose magnitudes and signs indicate the strength and direction of *local sensitivity* of metric F^j to small perturbations in each of the parameters, where

$$dF^j/d\theta_i = \frac{-2}{T}\sum_{t=1}^{T}\left(fZ_t - fY_t(\Theta^j)\right)\cdot\left.\frac{df}{dY_t}\right|_{\Theta^j}\cdot\left.\frac{dY_t}{d\theta_i}\right|_{\Theta^j} \tag{3a}$$

which can be written as

$$dF^j/d\theta_i = \frac{-2}{T}\sum_{t=1}^{T} r_t(\Theta^j)\cdot\beta_t(\Theta^j)\cdot\left.\frac{dY_t}{d\theta_i}\right|_{\Theta^j} \tag{3b}$$

Here $r_t(\Theta^j) = fZ_t - fY_t(\Theta^j)$ is the residual (total error), the term $\beta_t(\Theta^j) = \left.\frac{df}{dY_t}\right|_{\Theta^j}$ accounts for the transformation operator, and the sensitivity coefficient $\left.\frac{dY_t}{d\theta_i}\right|_{\Theta^j}$ quantifies the manner in which the model generated output Y_t responds to small changes in parameter θ_i at time step t, where all of these terms are evaluated at location Θ^j. From equations (3a) and (3b), we see that the local partial derivative $dF^j/d\theta_i$ of the metric is constructed as an *importance-weighted* average of the time series of sensitivity coefficients $\left.\frac{dY_t}{d\theta_i}\right|_{\Theta^j}$, where the importance of each term is determined by the multiplying factors $r_t(\Theta^j)$ and $\beta_t(\Theta^j)$. For example, if no transformation is applied, then we have $\beta_t = 1$, but in the case that a natural log transformation is applied we have $\beta_t = 1/Y_t$ and so the sensitivity terms associated with larger model outputs are de-emphasized by dividing by Y_t.

The critical issue, here, is that the importance of each local sensitivity coefficient term $\left.\frac{dY_t}{d\theta_i}\right|_{\Theta^j}$ in the time series is determined by the combined effects of the residual term $r_t(\Theta^j)$ that represents the *goodness of model fit* at that time step and the term $\beta_t(\Theta^j)$ that depends on the nature of the transformation function. Accordingly, the sensitivity rankings of the model parameters will be significantly affected by the specific manner in which the model tracks, or does not track, the observed data and by the particular choice of form of performance metric. Note also that errors in the observed data Z_t can affect the results.

While it is clear that the size and sign of the model residual term $r_t(\Theta^j)$ is relevant to a model *calibration* exercise, where the goal is to find parameter values that optimize the performance metric, the relevance of this term to an assessment of the sensitivity of the model outputs to variations in the parameters seems somewhat obscure. In fact, as shown by equations (3a) and (3b), if our analysis of parameter influence/importance is based on evaluating how a performance metric is affected by parameter changes, the time steps and parameter locations at which the model fits the data well (i.e., where $r_t(\Theta^j) \sim 0.0$) will, rather strangely, not influence the analysis. Counterintuitively, the results will instead be *biased* to represent time steps and parameter locations where the model performance is not very good (i.e., where $r_t(\Theta^j)$ is far from zero). There seems little, in general, to commend such an approach. In contrast, it makes much more sense to treat each of the $\left.\frac{dY_t}{d\theta_i}\right|_{\Theta^j}$ terms in the time series as being important, regardless of whether the value of $Y_t(\Theta^j)$ is close to the target value Z_t or not. We develop this insight further in the rest of this paper.

1.5. Terminology, Objectives, and Scope

The objective of this paper is to take a fresh look at the problem of parameter sensitivity analysis for DESMs. To begin, we suggest using a more correct terminology—here it is not actually the parameters that are sensitive (as in the inverse problem of parameter identifiability analysis) but instead the model behaviors (state and output trajectories) that are sensitive to the changes in the parameters. Accordingly, the GSA

procedure developed here will be referred to as a strategy for establishing (relative) parameter *importance* (not parameter sensitivity).

Motivated by the above discussion, we seek a GSA approach to establishing parameter importance that is based in a theoretically sound conceptual and mathematical development that: (a) recognizes the dynamical state-space structural form of DESMs and (b) clearly measures the extent to which different parameters exert stronger (or weaker) controls on the models' behavior. Accordingly, the approach developed in this paper will follow Method 4 (see section 1.2) and

1. Not require that system output data be available.
2. Be based explicitly on properties of the time series of sensitivity coefficients $\frac{dY_t}{d\theta_i}\Big|_{\Theta^j}$, as these are what contain the required information about sensitivity of the model outputs to parameter perturbations.

As a consequence, the parameter importance indices and rankings developed in this paper will not depend on the choice of metric and/or system response (observation) data used to characterize model performance or on their associated uncertainties or errors. However, a generalization of this approach to accommodate alternative choices of $\boldsymbol{R}$ and/or different basic ways of computing the sensitivity coefficient (section 1.2) will be presented in a subsequent paper.

Section 2 presents the development of our approach to GSA of DESMs. This includes assessment of both:

1. Total period time-aggregate parameter importance; this provides summary information on the controls of model behavior and is useful for selecting parameters to be varied/fixed during model calibration.
2. Time-varying parameter importance at various temporal scales; this provides information useful for study of model process representations and for evaluating effects of variability in system inputs.

Our discussion will also highlight the importance and role of various decisions that must be made by the user when implementing any such method for parameter importance analysis; this is important because results will necessarily depend on context and on the manner in which the analytical problem is set up and solved.

Next, section 3 demonstrates a practical real-data application of our GSA method to evaluate time-aggregate and time-varying parameter importance for the 10-parameter HBV-SASK model applied to the Oldman basin in Canada. The results are found to be clearly consistent with system and process understanding and efficiency appears to be high, with relatively few parameter location samples required to obtain highly stable parameter importance rankings. For illustration, we compare our results with the *Morris* and *Sobol'* approaches applied using performance metrics. In section 4 we summarize and discuss our findings and conclusions and present a discussion of challenges encountered and future directions to be pursued.

2. Global Sensitivity Analysis of Dynamical Models from First Principles

For simplicity of notation/discussion, the following mathematical development will be restricted to that of a specific model output $\boldsymbol{Y^k}(\boldsymbol{\Theta})$ of the set of $\boldsymbol{D_Y}$ outputs $\left\{\boldsymbol{Y^1}(\boldsymbol{\Theta}), \ldots, \boldsymbol{Y^{D_Y}}(\boldsymbol{\Theta})\right\}$ generated by the model. When the number of model outputs $\boldsymbol{D_Y} > 1$, the same mathematical treatment should be applied to each output. Obviously, and as will be demonstrated in section 3, overall assessments regarding parameter importance and importance rankings must necessarily take into account the multi-output nature of DESMs and the results obtained over all relevant model outputs of interest.

2.1. Characterization of Local Sensitivity (at a Point) in Terms of Derivatives

2.1.1. The Local Sensitivity Matrix

The classical, calculitic, way to characterize sensitivity of $\boldsymbol{Y^k}(\boldsymbol{\Theta})$ to changes in $\boldsymbol{\Theta}$ is in terms of the matrix of mathematical derivatives $\nabla_{\boldsymbol{\Theta}}\boldsymbol{Y^k}$. The values of these derivatives will (in general) vary with the location of $\boldsymbol{\Theta}$, which can be anywhere within the feasible space $\boldsymbol{\Phi_\Theta}$. Accordingly, we will use the notation $\boldsymbol{\Theta^j} = \left\{\theta_1^j, \ldots, \theta_{N_\theta}^j\right\}$, where the index $\boldsymbol{j}$ indicates a particular location (set of values for the parameters $\boldsymbol{\Theta}$) within $\boldsymbol{\Phi_\Theta}$, and use the indexed notation $\nabla_{\boldsymbol{\Theta}}^j\boldsymbol{Y^k}$ to refer to the matrix of derivatives computed at location $\boldsymbol{\Theta^j}$ in the parameter space.

Since there are $\boldsymbol{N_\theta}$ parameters and $\boldsymbol{T}$ time steps, the matrix $\nabla_{\boldsymbol{\Theta}}^j\boldsymbol{Y^k}$ is two-dimensional and consists of $\boldsymbol{N_\theta}$ rows (one for each parameter) and $\boldsymbol{T}$ columns (one for each time step):

$$\nabla^{j}_{\Theta}\boldsymbol{Y}^{k} = \begin{bmatrix} dY^{k}_{1}/d\theta^{j}_{1} & \cdots & dY^{k}_{T}/d\theta^{j}_{1} \\ \vdots & \ddots & \vdots \\ dY^{k}_{1}/d\theta^{j}_{N_\theta} & \cdots & dY^{k}_{T}/d\theta^{j}_{N_\theta} \end{bmatrix} \quad (4)$$

We refer to this as the *local sensitivity matrix* associated with output $\boldsymbol{Y^k}$. Since there are $\boldsymbol{D_Y}$ outputs, we will have a total of $\boldsymbol{D_Y}$ such matrices $\left\{\nabla^{j}_{\Theta}\boldsymbol{Y^1}, \ldots, \nabla^{j}_{\Theta}\boldsymbol{Y^{D_Y}}\right\}$.

Note that each row consists of a time series of derivatives $\boldsymbol{dY^k_t/d\theta^j_i}$ ($\boldsymbol{t = 1, \ldots, T}$) associated with a specific parameter (indexed by $\boldsymbol{i}$) at location $\boldsymbol{j}$ in the feasible parameter space for a specific output $\boldsymbol{Y^k}$. Accordingly, each row contains the information necessary to characterize the local sensitivity (at location $\boldsymbol{j}$) of the time series of model output values $\boldsymbol{Y^k}=\left\{\boldsymbol{Y^k_1}, \ldots, \boldsymbol{Y^k_T}\right\}$ to local (small) variations in each of the parameters $\boldsymbol{\Theta}=\{\theta_1, \ldots, \theta_{N_\theta}\}$.

So, to characterize the relative sensitivity of $\boldsymbol{Y^k}$ with respect to each parameter, we must compare the properties of the rows. Since larger derivative values $\boldsymbol{dY^k_t/d\theta^j_i}$ indicate larger local sensitivities, we must quantify (in some manner) the relative *sizes* of the rows.

2.1.2. Appropriate Scaling of Parameters

Since each row of $\nabla^{j}_{\Theta}\boldsymbol{Y^k}$ is a *vector* consisting of $\boldsymbol{t = 1, \ldots, T}$ values, we can use some summary property of this vector as a measure of its relative size. But, before discussing such measures, we should point out that any such analysis can depend (be conditional) on the a priori selection of an appropriate choice of scaling. This may be particularly true if a finite difference approach is used to estimate the derivatives used to populate $\nabla^{j}_{\Theta}\boldsymbol{Y^k}$.

In particular, one must pay special attention when the feasible range of a particular model parameter (say $\boldsymbol{\theta_j}$) varies over orders of magnitude (e.g., hydraulic conductivity). In such cases, it may be appropriate to transform that parameter by taking the logarithm $\boldsymbol{\theta_i^{Trans}}=\log_{10}(\theta_i)$ so as to not obtain inappropriately biased indications of parameter sensitivity.

Finally, a common approach is to scale such that variations of each parameter between its minimum and maximum values in $\boldsymbol{\Phi_\Theta}$ correspond to range **[0,1]**. Hereafter, we will adopt this convention and note that changes in scaling can alter the interpretations drawn.

2.1.3. Appropriate Scaling of Outputs

Similarly, it may be necessary to appropriately scale the outputs or to treat each output separately so that their relative sensitivities are not directly compared. However, even when dealing with each model output separately, one must pay special attention when the model output varies over orders of magnitude. For example, streamflow can vary from very low values during recession/dry periods to very high values during intense rainstorm events. If this situation is left untreated, there can be a tendency for parameters that govern the generation of high streamflow values during rainstorm events to *appear* to be more sensitive than parameters that govern the generation of low streamflow values during long recession periods.

In cases such as this, it may be appropriate to transform the output values so as to not obtain inappropriately biased indications of parameter sensitivity. For example, for streamflow it may be useful to perform a logarithmic or power transformation (Box & Cox, 1974) before performing the sensitivity analysis.

Hereafter, we will assume that a meaningful scaling of distances in the parameter and output spaces, if needed, has been properly established.

2.1.4. Aggregate Measures of Local Sensitivity

The $\boldsymbol{i}$th row of $\nabla^{j}_{\Theta}\boldsymbol{Y^k}$ consists of a time series of values $\nabla^{j}_{\theta_i}\boldsymbol{Y^k}=\left\{\boldsymbol{dY^k_1/d\theta^j_i}, \ldots, \boldsymbol{dY^k_T/d\theta^j_i}\right\}$ (see examples in first column of Figure 2). Although the derivative values are not serially independent, the frequency distribution $\boldsymbol{p^j_{1:T}}\left(\boldsymbol{dY^k_t/d\theta_i}\right)$ of these values can be thought of as characterizing the time-aggregated sensitivity properties of output $\boldsymbol{Y^k}$ with respect to parameter $\boldsymbol{\theta_i}$ at location $\boldsymbol{j}$ (see second column of Figure 2); for convenience, all of the frequency distributions shown in this paper have been normalized such that the total frequency sums to one. Note that derivatives can be either positive or negative (or zero). So when discussing sensitivity, if we are primarily concerned with the magnitude (strength) of change in $\boldsymbol{Y^k}$ to local (small) perturbations of $\boldsymbol{\theta_i}$,

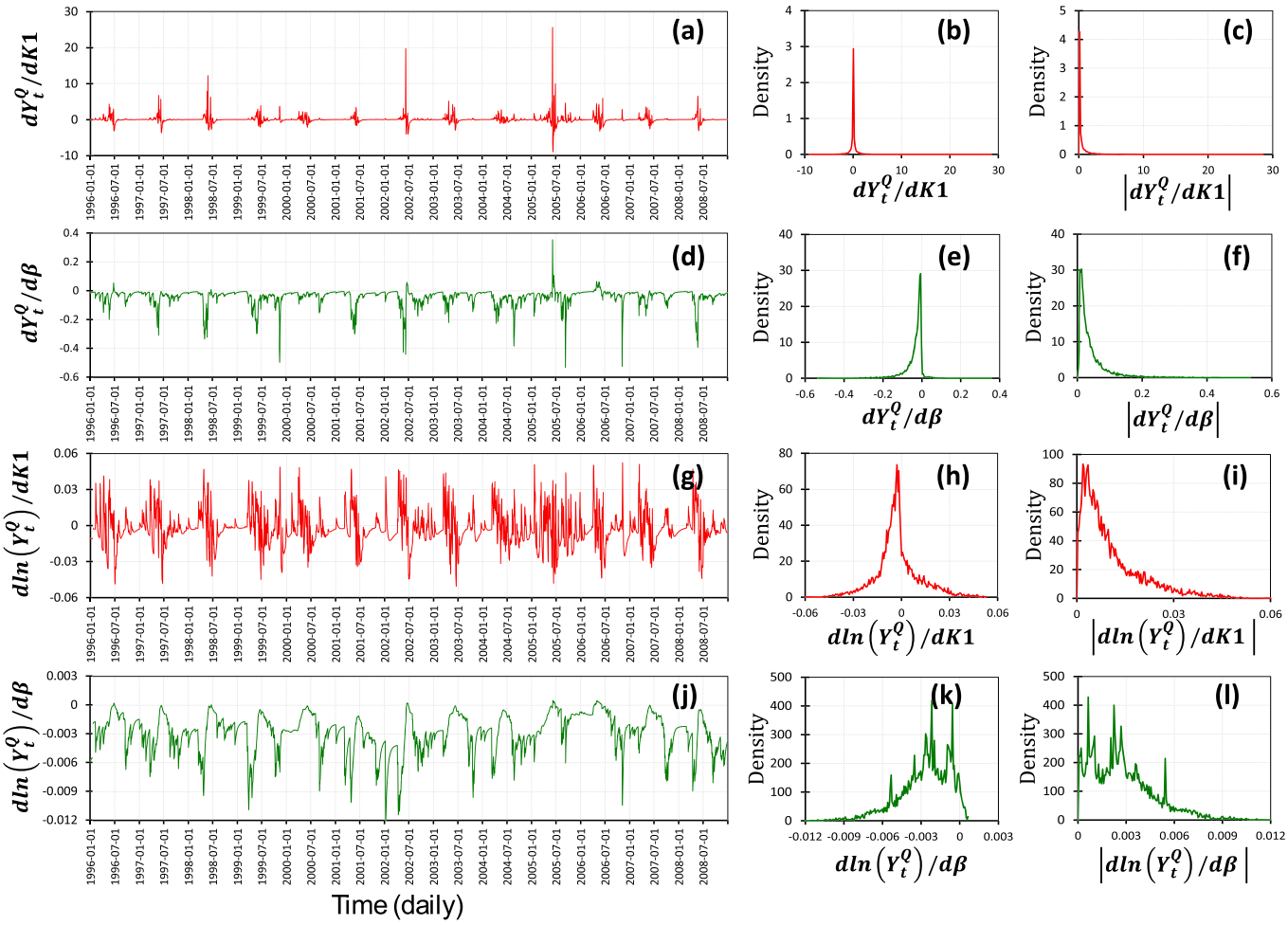

Figure 2. Example plots showing time series of sensitivity coefficients (partial derivatives) of streamflow Y_t^Q with respect to parameters ***K1*** and ***β*** (both known to be moderately influential) of the HBV-SASK model, when applied to the Oldman basin (see section 3): (a, d, g, and j) Time series of $dY_t^Q/dK1$, $dY_t^Q/d\beta$, $dln\left(Y_t^Q\right)/dK1$, and $dln\left(Y_t^Q\right)/d\beta$ respectively; (b, e, h, and k) corresponding frequency distributions of the derivatives; (c, f, i, and l) corresponding frequency distributions of the absolute values of partial derivatives.

we can focus on the properties of the frequency distributions $p_{1:T}^j\left(\left|dY_t^k/d\theta_i\right|\right)$ of *absolute* values (see third column of Figure 2).

With this perspective, distributions $p_{1:T}^j\left(\left|dY_t^k/d\theta_i\right|\right)$ that are spread out further to the right along the $|dY/d\theta|$ axis correspond to parameters θ_i to which the model outputs are *more frequently strongly sensitive* (see Figure 3). This requires us to compare distributions, a somewhat complicated task. However, a simple way to characterize, in summary fashion, the overall relative strengths of local sensitivities (at location ***j***) of Y^k with respect to θ_i would be to use some appropriate statistical measure (or measures) of the distribution of frequencies of $\left|dY_t^k/d\theta_i\right|$ obtained.

In selecting such a measure (or measures), we should take into consideration the typical shapes that can be expected for the frequency distributions of partial derivatives $dY_t^k/d\theta_i$. In general, these derivatives can range from large negative values (indicating decreasing Y_t^k when θ_i is increased) to large positive values (indicating increasing Y_t^k when θ_i is increased). Further, given that environmental systems often spend considerable periods of time in relaxation or nondriven mode, during which they are not being strongly stimulated by system drivers, we are likely to see a large frequency of partial derivative values that are zero or close to zero. Examples of such local (at a point) frequency distributions for the HBV-SASK hydrologic model can be seen in Figure 2 (and also in Figures 5a and 5b for *global* frequency distributions).

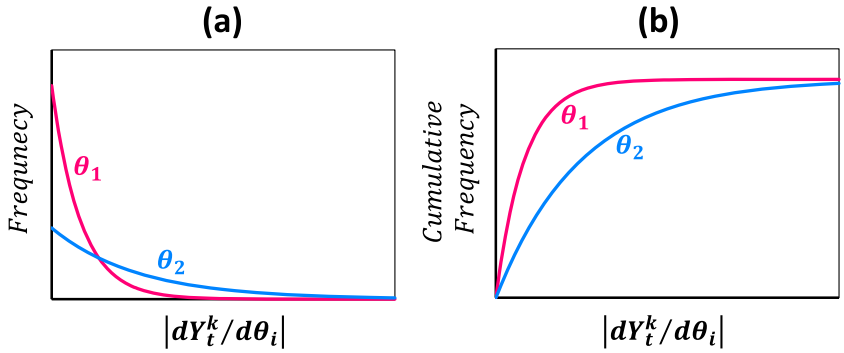

Figure 3. Cartoon illustrating typical frequency distributions constructed from absolute values of the time series of partial derivatives of a model output Y_t^Q with respect to two DESM model parameters θ_1 and θ_2. Here parameter θ_2 has a greater frequency of larger $|dY_t^k/d\theta_i|$ values than does parameter θ_1, indicating a stronger time-aggregate sensitivity of output Y^k to parameter θ_2.

Accordingly, we can expect the frequency distributions $p_{1:T}^j(|dY_t^k/d\theta_i|)$ of absolute values of the partial derivatives to be *exponential-like*, with modes at or close to zero and frequencies decaying toward the right. In such cases, we can use some measure of central tendency $C_i^k(j)$, such as the mean or median, to summarize the properties of each distribution. Alternatively, we could use some measure of the dispersion $D_i^k(j)$ of each distribution, such as its standard deviation, entropy, or distance between zero and some upper quantile (say 97.5%), thereby measuring how far the distribution spreads out toward larger values. However, for exponential-like distributions, we can expect the measures of central tendency $C_i^k(j)$ and dispersion $D_i^k(j)$ to be quite strongly correlated. For example, in the case of perfectly exponential distributions we have exact equivalence between the standard deviation and mean (i.e., $\sigma_i^k(j)=\mu_i^k(j)$), the median $\mathbf{med}_i^k(j)=\mu_i^k(j)\cdot\ln(2)$ and the entropy $H_i^k(j)=1+\ln(\mu_i^k(j))$.

Alternatively, one might consider using the Euclidian length $L_i^k(j)=\sqrt{\sum_{t=1}^{T}(dY_t^k/d\theta_i)^2}$ of the vector $\nabla_{\theta_i}^j Y^k$ as the measure of its relative size. However, simple algebraic manipulation shows that $L_i^k(j)=\sqrt{T}\cdot\sqrt{[\mu_i^k(j)]^2+[\sigma_i^k(j)]^2}$, where μ is the mean and σ is the standard deviation of $p_{1:T}^j(|dY_t^k/d\theta_i|)$, and so for an exponential distribution the result is effectively the same (i.e., since $\sigma_i^k(j)=\mu_i^k(j)$, we have $L_i^k(j)=\sqrt{2T}\cdot\mu_i^k(j)$).

Accordingly, in the remaining presentation we will adopt the mean $\mu_i^k(j)$ of the distribution of absolute partial derivatives as our time-aggregate measure of the strength of local sensitivity. In doing so, we will be careful to verify that the shape of the corresponding frequency distribution is indeed exponential-like. More generally, appropriate application of the methods developed in this paper will necessitate that the aggregate statistical measure(s) selected to indicate sensitivity strength be carefully chosen based on the actual nature of the distributions obtained in a given situation. It is possible that in some situations the entropy of the distribution or the distance between zero and some upper quantile (say 97.5%) will be a more appropriate measure.

2.2. Characterization of Global Sensitivity

2.2.1. The Global Sensitivity Matrix

Any measure of sensitivity computed at a particular location j in the feasible parameter space Φ_Θ is specific to only that location. Therefore, to properly characterize the sensitivity of a given model output Y^k across the feasible parameter space (i.e., in some global sense), it becomes necessary to conduct an evaluation of how the derivative values vary across that entire space.

Since, in general, the feasible parameter space is a continuum consisting of an uncountable number of distinct parameter locations, we will base our analysis on a study of properties of the derivatives evaluated at a *sufficiently-well sampled* set of N_{Pts} representative locations throughout Φ_Θ. Based on these *sampled* values, we derive statistical estimates of various measures of global relative sensitivity of the model outputs to

perturbations of the parameters (i.e., measures of parameter *importance*). Of course, since any such results will be subject to sampling variability, we must take that variability into account when evaluating the results.

So now for each output $\boldsymbol{Y^k}$, instead of a two-dimensional local sensitivity matrix of dimension $\boldsymbol{N_\theta \times T}$, we have a three-dimensional *global sensitivity matrix* $\nabla_{\boldsymbol{\Theta}}\boldsymbol{Y^k}$ of dimension $\boldsymbol{N_{Pts} \times N_\theta \times T}$, where $\boldsymbol{N_{Pts}}$ is the number of sample locations, $\boldsymbol{N_\theta}$ is the number of parameters, and $\boldsymbol{T}$ is the number of time steps. This three-dimensional matrix can be thought of as a set of $\boldsymbol{N_\theta}$ two-dimensional ($\boldsymbol{N_{Pts} \times T}$) global sensitivity matrices $\nabla_{\boldsymbol{\Theta}}\boldsymbol{Y^k}(\boldsymbol{i})$, one for each parameter $\boldsymbol{\theta_i}$:

$$\nabla_{\boldsymbol{\Theta}}\boldsymbol{Y^k}(\boldsymbol{i}) = \begin{bmatrix} \boldsymbol{dY_1^k/d\theta_i^1} & \cdots & \boldsymbol{dY_T^k/d\theta_i^1} \\ \vdots & \ddots & \vdots \\ \boldsymbol{dY_1^k/d\theta_i^{N_{Pts}}} & \cdots & \boldsymbol{dY_T^k/d\theta_i^{N_{Pts}}} \end{bmatrix} \tag{5}$$

where the columns correspond to time and each row corresponds to one of the $\boldsymbol{N_{Pts}}$ locations sampled across the feasible parameter space (we now have both an ensemble of model trajectory time series and their associated derivative time series).

2.2.2. Total Period Time-Aggregate Global Sensitivity

To characterize the relative global sensitivity of output $\boldsymbol{Y^k}$ to local perturbations in the parameter $\boldsymbol{\theta_i}$, we must develop some aggregate measure (or measures) that quantifies the relative sizes of the $\nabla_{\boldsymbol{\Theta}}\boldsymbol{Y^k}(\boldsymbol{i})$ matrices. Following the method proposed in section 2.1.4, this can be done by analyzing the properties of the frequency distributions $\boldsymbol{p_{1:T}^{1:N_{Pts}}}\left(\boldsymbol{dY_t^k/d\theta_i}\right)$ of the $\boldsymbol{N_{Pts} * T}$ component values of each $\nabla_{\boldsymbol{\Theta}}\boldsymbol{Y^k}(\boldsymbol{i})$ matrix or more specifically the properties of the distributions $\boldsymbol{p_{1:T}^{1:N_{Pts}}}\left(\left|\boldsymbol{dY_t^k/d\theta_i}\right|\right)$ of their absolute values.

As discussed before, proper consideration must be given to appropriate scaling of the parameters (section 2.1.3) and the outputs (section 2.1.4). Having done so, aggregate indicators of global sensitivity can be constructed based on statistical measures that summarize the properties of the global distributions of sensitivity coefficients $\boldsymbol{p_{1:T}^{1:N_{Pts}}}\left(\boldsymbol{dY_t^k/d\theta_i}\right)$ for each parameter. Following section 2.1.4, for exponential-like distributions we can use the means $\boldsymbol{\mu_i^k}$ of the distributions of globally sampled absolute partial derivatives as our time-aggregate measures of the strengths of global sensitivity associated with each parameter, while for distributions that are not sufficiently exponential-like other statistical measures may be more appropriate.

2.2.3. Time-Varying Global Sensitivity: Consideration of Temporal Dynamics

The previous discussion has focused on measures of (local and global) sensitivity that are aggregate properties of the entire model simulation time period. In general, however, such aggregation over the entire time period can obscure the important fact that, given the complex nonlinear nature of the dynamics of environmental systems, each of the outputs $\boldsymbol{Y^k}$ generated by a model of the system will typically show varying degrees (strengths) of sensitivity to various parameters across time.

For example, certain model outputs may display significant sensitivity to specific parameters only during time periods when the ***PPEs*** associated with those parameters are significantly activated by the system inputs (e.g., Gupta & Sorooshian, 1985; Guse et al., 2016; Wagener et al., 2003). So in a catchment model, the parameters associated with overland flow generation will only be activated during periods when the system is driven by rainfall and for some time thereafter while rainwater remains ponded on the surface.

To track this kind of behavior, we can partition the three-dimensional ($\boldsymbol{N_{Pts} \times N_\theta \times T}$) global sensitivity matrix $\nabla_{\boldsymbol{\Theta}}\boldsymbol{Y^k}$ into $\boldsymbol{T}$ two-dimensional ($\boldsymbol{N_\theta \times N_{Pts}}$) *time-specific* global sensitivity matrices $\nabla_{\boldsymbol{\Theta}}\boldsymbol{Y^k}(\boldsymbol{t})$, one for each time $\boldsymbol{t = 1, \ldots, T}$, where

$$\nabla_{\boldsymbol{\Theta}}\boldsymbol{Y^k}(\boldsymbol{t}) = \begin{bmatrix} \boldsymbol{dY_t^k/d\theta_1^1} & \cdots & \boldsymbol{dY_t^k/d\theta_1^{N_{Pts}}} \\ \vdots & \ddots & \vdots \\ \boldsymbol{dY_t^k/d\theta_{N_\theta}^1} & \cdots & \boldsymbol{dY_t^k/d\theta_{N_\theta}^{N_{Pts}}} \end{bmatrix} \tag{6}$$

such that each row corresponds to one of the $\boldsymbol{N_\theta}$ parameters and each column corresponds to one of the $\boldsymbol{N_{Pts}}$ locations sampled across the feasible parameter space (we now have a time series of $\nabla_{\boldsymbol{\Theta}}\boldsymbol{Y^k}(\boldsymbol{t})$ matrices).

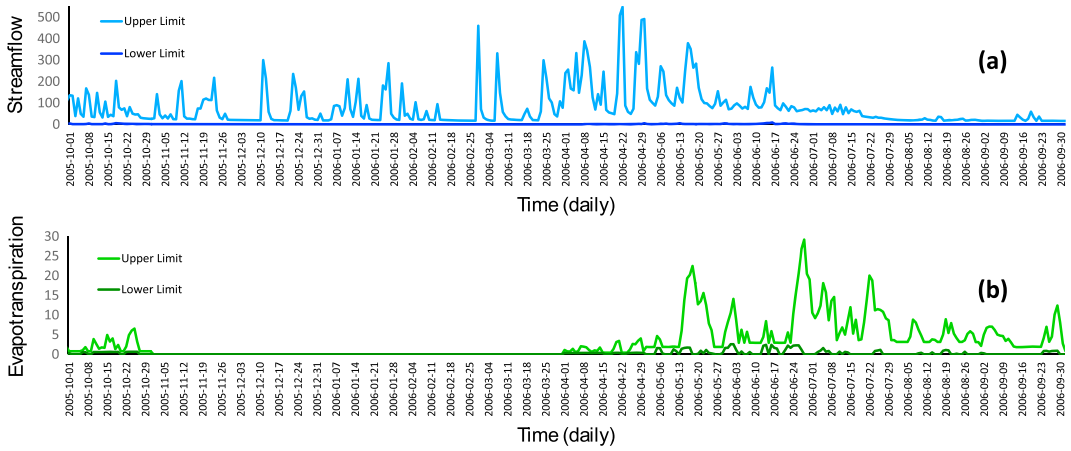

Figure 4. Time series plots illustrating the time varying spreads for an ensemble of simulated (a) streamflow and (b) evapotranspiration trajectories, for a representative period (October 2015 to September 2016), obtained by sampling N_{Pts} parameter locations across the feasible parameter space of the HBV-SASK model applied to the Oldman basin.

Consider now the vector $\nabla_{\theta_i} Y^k(t) = \{dY_t^k/d\theta_i^1, \ldots, dY_t^k/d\theta_i^{N_{Pts}}\}$ defined by the ith row of $\nabla_\Theta Y^k(t)$, whose elements correspond to *all* of the N_{Pts} values of the sensitivity of output Y^k to parameter θ_i sampled across the feasible space, at time t. Accordingly, we can characterize the relative *time-specific global* sensitivities of output Y^k to local perturbations in each of the parameters θ_i by analyzing the properties of the frequency distributions $p_t^{1:N_{Pts}}(dY_t^k/d\theta_i)$ of the N_{Pts} component values of each of the i rows of the matrix $\nabla_\Theta Y^k(t)$ for time **t**, or more specifically the properties of the distributions $p_t^{1:N_{Pts}}(|dY_t^k/d\theta_i|)$ of their absolute values.

Now we can examine how the degree (strength) of global sensitivity of output Y^k to each parameter varies with time. So distributions $p_t^{1:N_{Pts}}(|dY_t^k/d\theta_i|)$ that spread out further to the right along the $dY/d\theta$ axis correspond to parameters to which the model outputs are *more strongly globally sensitive* at a given time t, and we can characterize the *global sensitivity at time* t of output Y^k with respect to parameter θ_i via statistical measures that characterize the central tendency or dispersion of these global distributions. By plotting the time variation of those measures (e.g., of the means $\mu_i^k(t)$), we can track how the global sensitivities vary with time.

2.2.4. Consideration of the Signs of the Derivatives

In addition to the *strength* of sensitivity, the *direction* of sensitivity (whether the signs of the derivatives are positive or negative) of output Y^k to various parameters can also vary with time. Since, when analyzing the behavior of a model, information regarding whether a sensitivity value is dominantly positive or negative at any given time can be very useful, we propose that it is important to keep track of such information.

To do so, we can examine the properties of the frequency distribution $p_t^{1:N_{Pts}}(dY_t^k/d\theta_i)$ of values where the operation of taking their absolute values has *not* been applied. Now we can track how the strength of positive and negative sensitivity varies with time by separately computing the statistical measures characterizing sensitivity strength (e.g., $\mu_i^k(t,+)$ and $\mu_i^k(t,-)$) associated with the probability masses on either side of the origin.

2.2.5. Temporal Reweighting or Grouping and When It Might Be Necessary or Useful

When we generate an ensemble of model output trajectories by sampling a set of N_{Pts} parameter locations across the feasible parameter space, it can happen that the spread of the output ensemble varies quite considerably with time. For example, Figure 4a illustrates the streamflow value spread that can be obtained when sampling the feasible parameter space of a conceptual rainfall-runoff model. In this case, the larger spread during rainfall-driven flood events is caused by the (nonlinear) sensitivity of model-generated streamflow to the larger degree of activation of the system by the input driver (rainfall), while the smaller spread and its progressive diminishment during recession events is due to the strongly dispersive nature of the catchment system when rainfall is not present.

When this happens, the magnitudes of the sensitivity indices (derivatives) can vary quite considerably with time, purely due to the fact that these sensitivities depend strongly on the variations in magnitude of the

input driver activation. So the possibility arises that at certain times the model outputs *seem* to be relatively insensitive to perturbations in a given parameter, when, in fact, this parameter is actually exerting a strong relative (to other parameters) influence at those times, but the impact of this sensitivity is masked by the temporal variations in model output scaling (i.e., the varying spread of the output ensemble). Note that the spread of an output ensemble at a given time directly depends (in some manner) on the overall sensitivity of the output to parameter perturbation throughout its feasible space.

In such cases, it can be useful (even important) to perform temporal reweighting or grouping to try and minimize the influence of such behavior on the overall results of the analysis. We can think of at least three ways in which this might be done:

1. *Rescaling Model Outputs:* Before performing the time-varying GSA as discussed above, one can rescale the ensemble of model outputs at each time step so that the overall range of variation is more or less constant through time. For example, in conceptual rainfall-runoff modeling we could rescale the ensemble of streamflow values to a range of 0–1 at each time step. Alternatively, we can use a logarithmic or power transformation of streamflow as discussed in section 2.1.3, as a partial way of reducing the variability with time.
2. *Rescaling Sensitivity Indices*: Before computing the measures of *Time-Aggregate* or *Time-Varying* global sensitivity as discussed in sections 2.3.2–2.3.5, we can adjust the elements of the three-dimensional ($\boldsymbol{N_{Pts}} \times \boldsymbol{N_\theta} \times \boldsymbol{T}$) global sensitivity matrix $\nabla_{\boldsymbol{\Theta}}\boldsymbol{Y^k}$ such that at each time step $\boldsymbol{t}$, the ($\boldsymbol{N_\theta} \times \boldsymbol{N_{Pts}}$) submatrix of derivatives $\nabla_{\boldsymbol{\Theta}}\boldsymbol{Y^k}(\boldsymbol{t})$ (see equation (6)) is normalized to have the exact same size; that is, such that either the sum $\sum_{j=1}^{N_{Pts}}\sum_{i=1}^{N_\theta} \boldsymbol{dY_t^k}/\boldsymbol{d\theta_i^j}$ of all the elements in the matrix or the determinant $\boldsymbol{Det}\ \nabla_{\boldsymbol{\Theta}}\boldsymbol{Y^k}(\boldsymbol{t})$ of the matrix is some arbitrary constant value, say $\boldsymbol{K}$. The effect of this rescaling would be to make the overall *total sensitivity* constant with time, so that every time step contributes equally to the analysis.
3. *Grouping Time Steps*: One can partition the time series of outputs into different classes, segments, or periods of time that correspond to various features of interest in the output (such as periods when different system processes are dominant) and compute Time-Aggregate global sensitivity indices separately for each period.

In regard to grouping, at one extreme we could compute global sensitivity indices for the entire time period as discussed in section 2.2.2 (a common approach), while at the other extreme we could perform a separate time-specific sensitivity ranking for each time step as in section 2.2.3 (see Guse et al., 2016). However, somewhere in between these two extremes may prove to be a most informative and useful approach whereby the sensitivity ranking is conducted separately on sets of *interesting temporal features* associated with each output $\boldsymbol{Y^k}$ (where the number of features is much smaller than the number of time steps).

So, for example, one could group based on temporal features that correspond to separate (or grouped) consecutive features in the output time series. In catchment rainfall-runoff modeling such features could correspond to rising and falling limbs of the hydrograph, or to raining and nonraining (driven and nondriven) periods (Boyle et al., 2000; Vrugt et al., 2003) or to summer and winter periods, etc. An alternative approach could be to first extract characteristic diagnostic features (patterns) from the output time series (Gupta et al., 2008); in catchment rainfall-runoff modeling such features could correspond to different properties of the flow duration curve (mean level, slope, upper and lower quantiles, etc. (Yilmaz et al., 2008).

The important point is that just as there is a need to make careful choices about parameter scaling and about what properties of the distributions of derivatives are used to characterize sensitivity, the output attributes to be studied must also be properly contextualized for the results of the sensitivity analysis to be useful and meaningful.

2.3. Summary of our Global Sensitivity Matrix Approach to GSA of DESMs

We will call our approach to GSA of DESMs the Global Sensitivity Matrix (GSM) approach. The steps in this method are as follows:

1. Define the Problem:

Select the model outputs $\boldsymbol{Y^k}$ and parameters $\boldsymbol{\Theta}$ of interest. Apply appropriate transformations (e.g., log or power) to variables that vary over several orders of magnitude. Scale each (transformed) parameter to the range [$\boldsymbol{0}$, $\boldsymbol{1}$]. For outputs, any meaningful range can be used.

2. Select a Representative Statistical Sample of Parameter Locations:

Generate N_{Pts} base sample locations distributed uniformly across the feasible parameter space. Use Progressive Latin Hypercube Sampling (Sheikholeslami & Razavi, 2017) to maximize representativeness and minimize potential statistical bias; this facilitates increasing sample size in a stepwise manner to achieve statistical robustness (reliability) while controlling computational cost.

3. Construct the Global Sensitivity Matrix:

Run the model for each base sample location, generate the partial derivative time series $dY_t^k/d\theta_i$ for each output and parameter, and construct the three-dimensional ($N_{Pts} \times N_\theta \times T$) global sensitivity matrix $\nabla_\Theta Y^k$ for each output. Where analytical derivatives are not available, use finite difference approximation with perturbation step sizes of 1% (or 5%) of the parameter range. If desired, apply temporal weighting or grouping to adjust the global sensitivity matrix $\nabla_\Theta Y^k$ (see section 2.2.5).

4. Conduct the Time-Aggregate Parameter Importance Analysis:

Compute the total period *Time-Aggregate Parameter Importance Indices* (see section 2.2.2) and rank/group the parameters in terms of overall relative importance.

5. Conduct the Time-Varying Parameter Importance Analysis:

Compute the time series of *Time-Varying Parameter Importance Indices* (see section 2.2.3). Do this both for the absolute derivatives $\left|dY_t^k/d\theta_i\right|$ (see section 2.2.4) and for the positively and negatively signed subsets of the derivatives (see section 2.2.4).

3. Illustrative Example

3.1. Study Site

We illustrate the GSM method by analyzing the sensitivity of the hydrologic response of the HBV-SASK conceptual rainfall-runoff model to perturbations of its 10 parameters, when applied to the Oldman (1,435-km^2) river basin in the Rocky Mountains of Alberta, Canada. Historical data are available for 1979–2008 (Figure 8 shows one typical year), from which we estimate average annual precipitation (rainfall + snowfall) to be 611 mm and average annual streamflow to be 11.7 m^3/s at gauge 05AA023 on the Oldman. The Oldman basin has a runoff ratio of ~0.42.

The method was also applied to HBV-SASK for the nearby Banff river basin (2,179 km^2, data available for 1950–2011, average annual streamflow = 38.6 m^3/s, runoff ratio ≈0.7). In general, parameter importance results for the two basins turned out to be quite similar, except in one important regard as discussed later. Accordingly, we present results mainly for the Oldman.

3.2. Conceptual Rainfall-Runoff Model

The HBV-SASK model (Figure 1) was coded at the University of Saskatchewan for educational purposes, based on an interpretation of the Hydrologiska Byråns Vattenbalansavdelning model (Lindström et al., 1997). The model has 10 parameters (Table 1) plus two additional coefficients (listed as parameters 11 and 12) whose values must be specified; in this study, the aforementioned coefficients were fixed at default values. In addition to the process parameterization equations listed in Figure 1, daily potential evapotranspiration (PET) is computed via the equation $\mathbf{PET}_t = (\mathbf{1} + \mathbf{ETF} \cdot (T_t - T_{mth})) \cdot \mathbf{PET}_{mth}$ where T_{mth} and $\mathbf{PET}_{mth}$ are the long-term average monthly temperature and potential evapotranspiration, respectively (both supplied as data), T_t is the temperature for the day, and **ETF** is a parameter (see Table 1). The feasible parameter space for the HBV-SASK model is defined via upper and lower bounds listed in Table 1. Parameters were scaled on the range [**0**, **1**]. Since the streamflow and evapotranspiration outputs do not vary over orders of magnitude, no output transformation was applied.

3.3. Time-Aggregate Parameter Importance Analysis

A base sample of N_{Pts} **= 10,000** random parameter locations was sampled uniformly from the 10-parameter feasible parameter space using Progressive Latin Hypercube Sampling. At each location, the time series of partial derivatives of model outputs (streamflow and evapotranspiration) with respect to the parameters were computed via finite difference approximation using a step size of 1% of the parameter range;

Table 1
Parameters of the HBV-SASK 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, that is, 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.

accordingly, the corresponding total number of model runs is $\boldsymbol{N_{Runs} = N_{Pts} \cdot (N_\theta + 1) = 110{,}000}$. We first look at the results obtained when all of the years in the available historical data period are lumped together. Then we take a look at the interannual variation.

3.3.1. Total Period Time-Aggregate Parameter Importance Analysis

Figure 5 shows cumulative CDFs of the actual, absolute, and time-normalized absolute values of the globally sampled partial derivatives for streamflow (left column) and evapotranspiration (right column), for the Oldman. Note that *x* axes in the second and third rows of plots are shown using a log scale. For time normalization, the sensitivity coefficients at each time step were adjusted such that their mean becomes 1.0 (second procedure in section 2.2.5). Clearly, the absolute value and time-normalized absolute value distributions tend to be exponential-like, although some cross each other in the tail regions where large derivative values occur with relatively low frequencies. Accordingly, for purposes of this demonstration, we will use the means μ_i^k of these distributions as our indices of sensitivity strength, and in Figure 6 we plot the parameter importance indices μ_i^Q for streamflow against the corresponding values μ_i^{ET} for evapotranspiration, for each of the parameters; the left subplot is for absolute derivatives and the right subplot is for time-normalized absolute derivatives.

For streamflow, the two parameters ***CO*** (base temperature melt factor) and ***TT*** (air temperature threshold) stand out as being considerably more important than the others, which makes sense given the snow-dominated nature of these basins. Interestingly, when time normalization is applied, an additional parameter—***K2*** (Slow reservoir drainage rate)—appears among the more important parameters, followed by ***FRAC*** (fraction of soil release entering fast reservoir), both of which control the nature of flow generated by drainage from the soil zone.

For evapotranspiration, the parameter ***ETF*** (evapotranspiration temperature-anomaly correction factor) is considerably more important than the others, followed by parameters ***CO*** (base temperature melt factor), ***FC*** (field capacity of soil), and ***LP*** (PET multiplier) showing moderate importance. The sensitivity of evapotranspiration to parameter ***CO*** makes sense, because this parameter acts to control the availability (supply) of water for evapotranspiration. Relatively low sensitivity is seen to parameters ***TT*** (rain/snow temperature threshold) and ***β*** (exponent in the nonlinear soil release equation). Finally, as it should be, sensitivity is zero to parameters ***FRAC***, ***K1***, ***K2***, and ***α***, all of which have to do with horizontal routing of water.

Note that if we were to perform the GSA based purely on sensitivity of streamflow to parameter changes, as is common in performance metric-based approaches because observed data regarding evapotranspiration is not available, we would miss the fact that ***ETF***, ***LP***, and ***FC*** are important parameters controlling the dynamical behavior of the model. Therefore, for a comprehensive understanding of parameter importance, it is important to apply a *multiobjective* (multi-output) approach to GSA (Rosolem et al., 2012). This cannot be achieved if our methodology depends on the availability of output flux data.

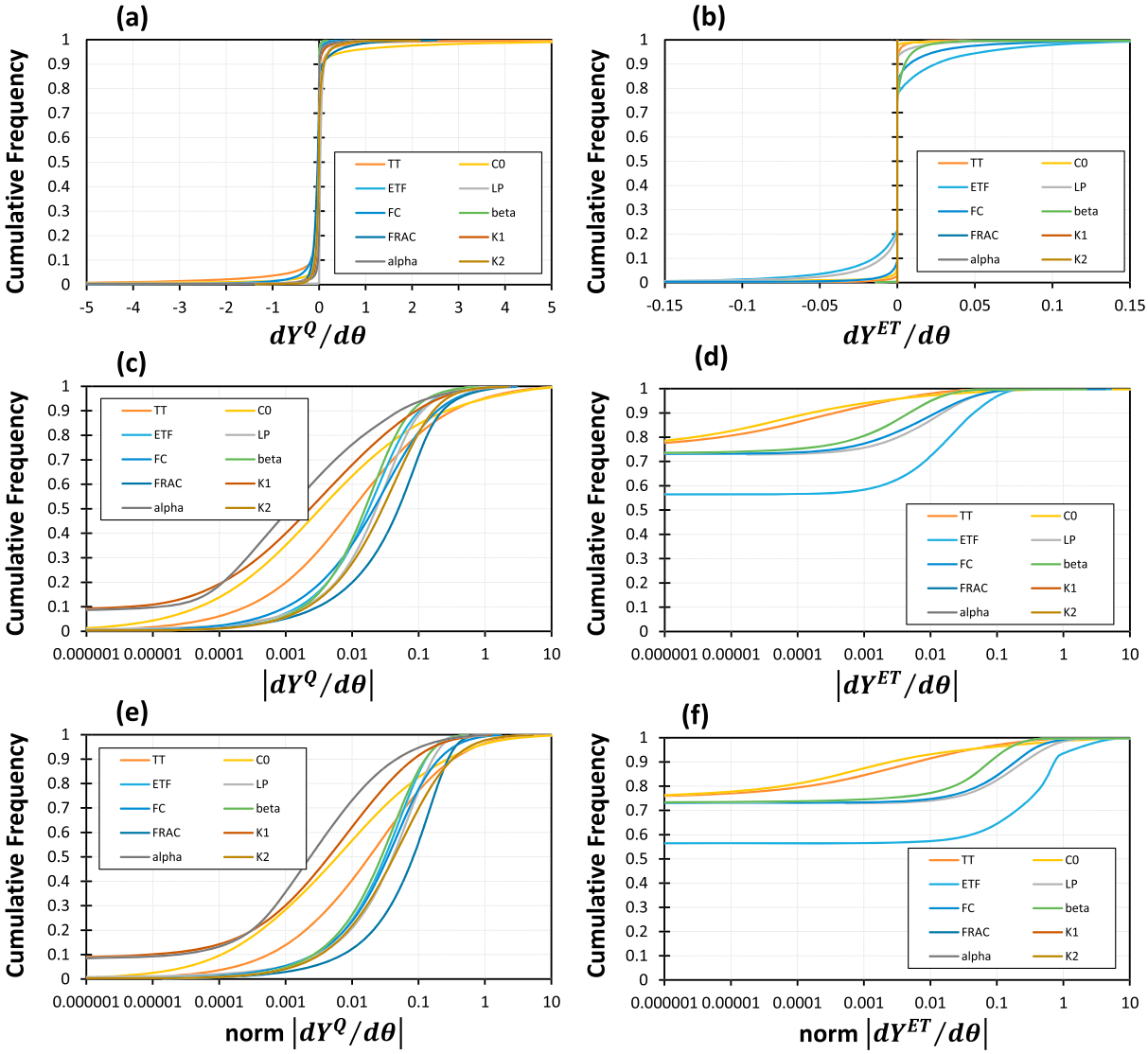

Figure 5. Plots showing total period *time-aggregate* cumulative frequency distribution functions of the globally sampled values of partial derivatives ($dY_t^k/d\theta_i$) obtained for the Oldman basin. Left column is for streamflow flux Y_t^Q and right column is for evapotranspiration flux Y_t^{ET}; (a and b) show actual partial derivatives, which can be both positive and negative; (c and d) show absolute values; (e and f) show time-normalized absolute values. Each colored line is for a different parameter.

Based on this analysis, we can assign the parameters ***CO***, ***ETF***, and ***TT*** to the category of *Strong Importance* and ***LP***, ***FC***, and ***FRAC*** to the category of *Moderate Importance* for this watershed, while, in a relative sense, the parameters α, β, ***K1***, and ***K2*** clearly exert the weakest controls on model behavior.

3.3.2. Annual Variability in Time-Aggregate GSA

In the previous section, we computed the parameter importance indices using the historical period 1982–2008 (the 3 years 1978–1981 were used as spin-up). However, the sensitivity of model response depends unavoidably on the nature and strength of the climatic forcing data that drives the system, and hence the model. Figure 7 (left column) shows the annual variability of *time-aggregate* parameter importance indices for streamflow and evapotranspiration. From subplot (b) it is clear that parameters ***CO*** and ***TT*** remain the most important parameters controlling streamflow, regardless of the annual-scale temporal variability in climatic system drivers. However, there is strong interannual variability of the importance indices of these parameters (between 0.15 and 0.45) due mainly to the large interannual variability in snowmelt. It is also interesting that the indices for the two parameters move in lockstep, and so it is debatable whether one or the other is more

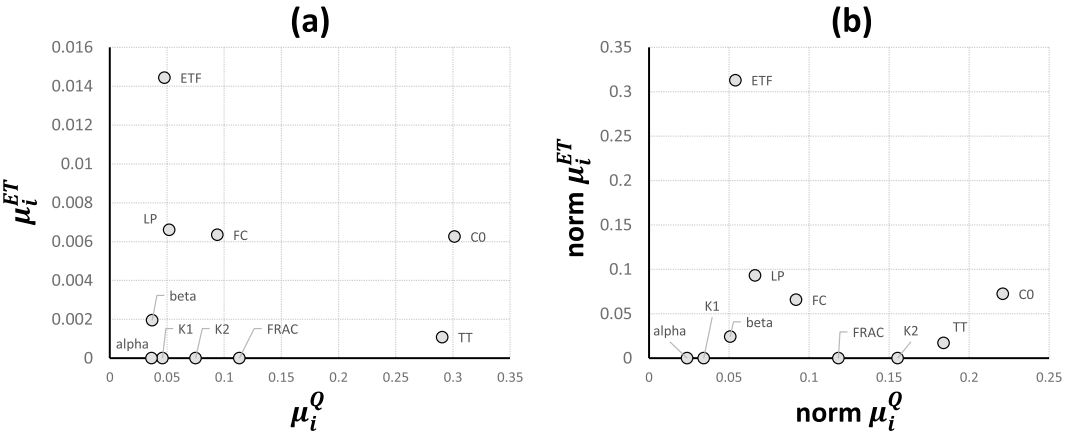

Figure 6. Plots showing total period *time-aggregate* GSA results for the Oldman basin. Parameter importance indices for streamflow flux $\boldsymbol{Y_t^Q}$, and evapotranspiration flux $\boldsymbol{Y_t^{ET}}$ are plotted on the *x* axis and *y* axis respectively; in each case, the indices shown are the *means* of the frequency distribution functions of absolute partial derivatives ($\left|\boldsymbol{dY_t^k/d\theta_i}\right|$ shown in Figure 5. The left plot is for derivative values, and the right plot is for time-normalized values.

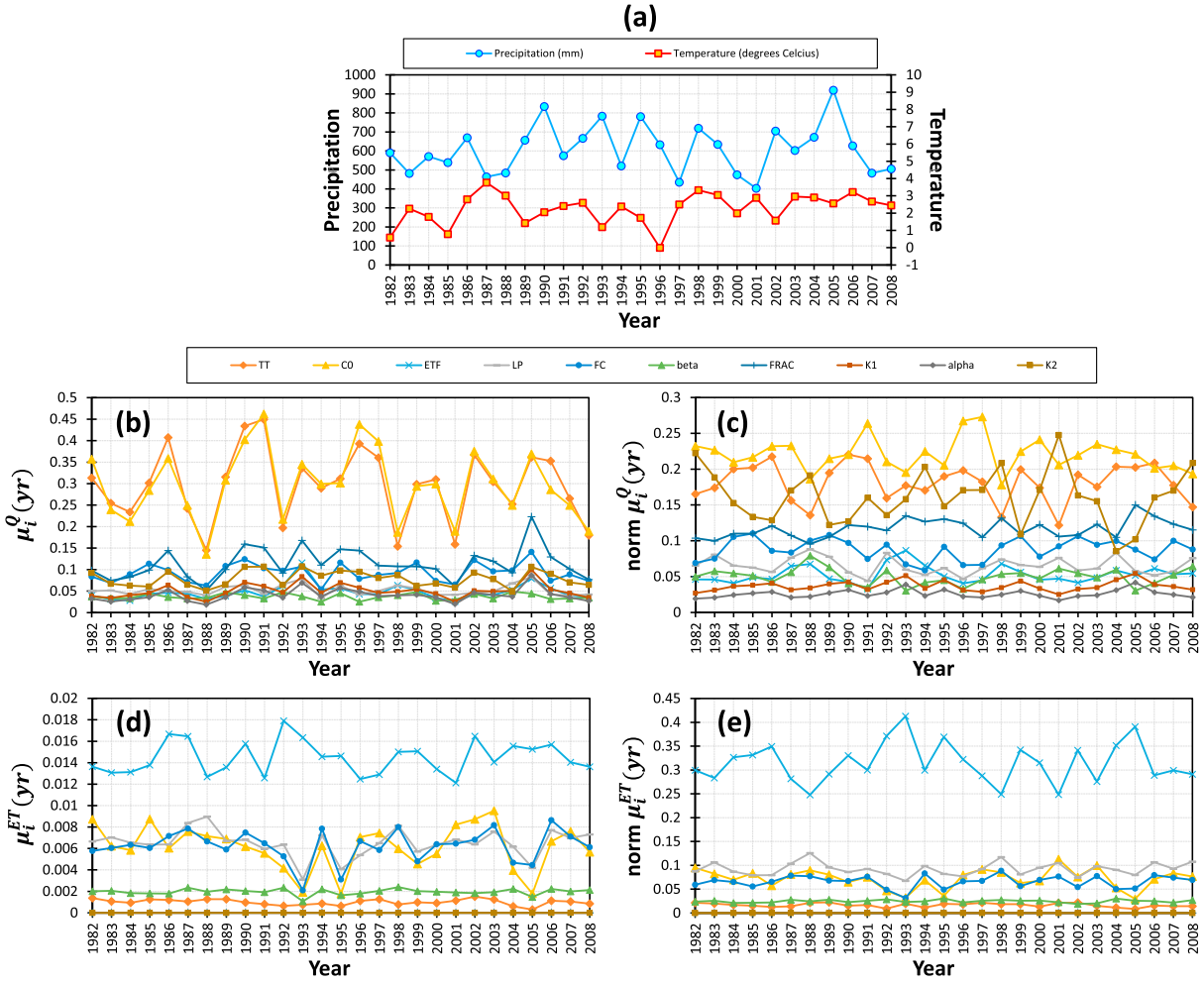

Figure 7. Plots showing dependence of GSM time-aggregate GSA results on annual variability in system drivers at the Oldman basin, where the parameter importance indices are computed separately for each year over the entire period. The second row is for streamflow $\boldsymbol{Y_t^Q}$, and the bottom row is for evapotranspiration $\boldsymbol{Y_t^{ET}}$. In the left column, the parameter importance index shown is the mean of the frequency distribution of absolute partial derivatives ($\left|\boldsymbol{dY_t^k/d\theta_i}\right|$. The right column shows corresponding results using time-normalized values. For reference, the top row shows the trajectories of mean annual precipitation and temperature.

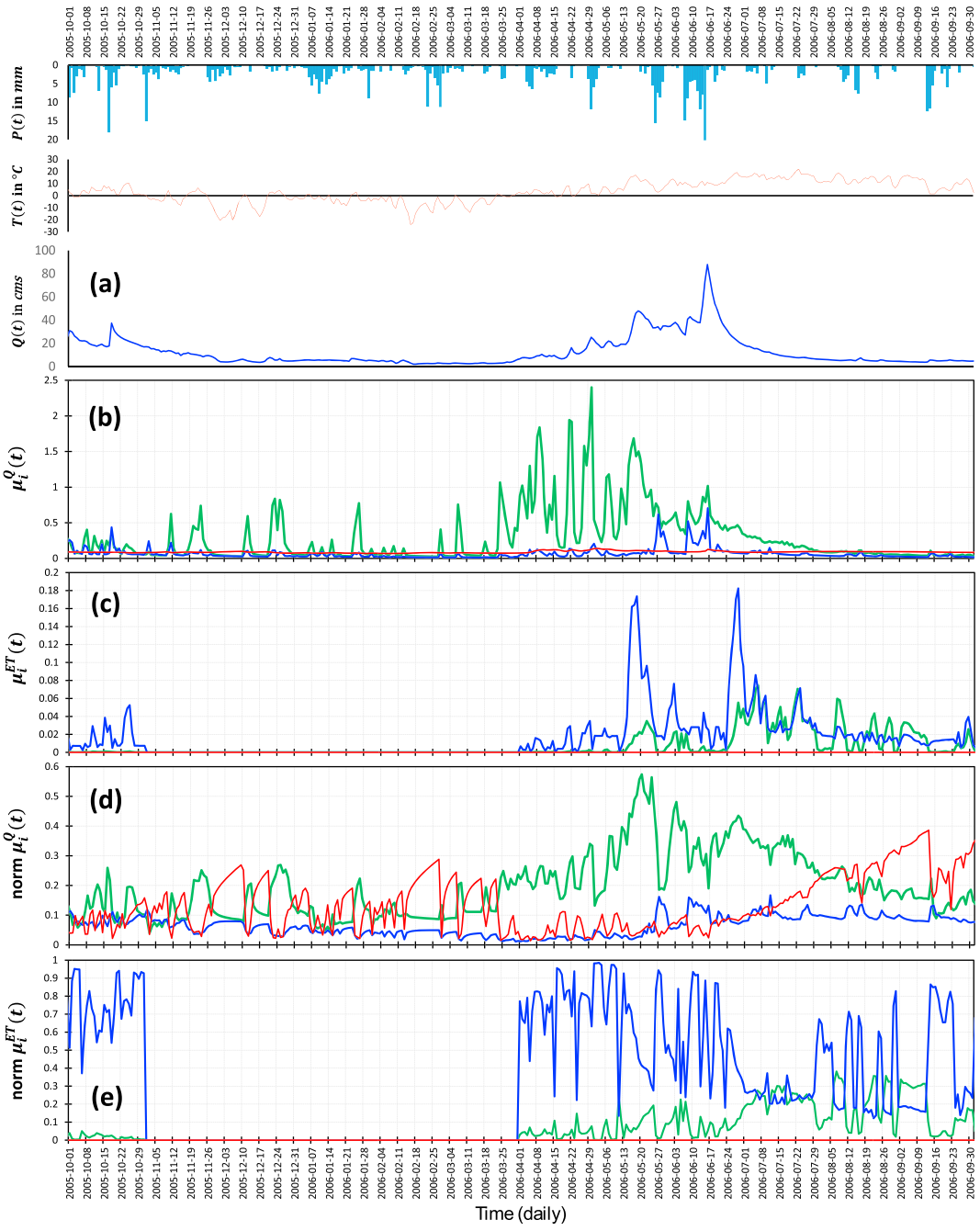

Figure 8. Plots showing GSM time-varying GSA for a representative period (October 2015 to September 2016) for the Oldman basin; (a) observed precipitation, temperature, and streamflow; (b and c) daily time series of absolute value parameter importance indices for streamflow and evapotranspiration; (d and e) similar plot showing *absolute time-normalized value* sensitivity indices. For illustration only three selected parameters are shown (green = CO, blue = ETF, and red = K2).

important. In contrast, the relative importance (ranking with respect to streamflow) of the other parameters tends to vary from year to year.

Similarly, from subplot (d), it is clear that parameter ***ETF*** remains the most important parameter controlling evapotranspiration, and in this case the interannual variability is much less. Meanwhile, the moderately influential parameters ***LP***, ***FC***, and ***CO*** seem to act as a group, moving in lockstep with each other. Finally, the low-importance parameters $\boldsymbol{\beta}$ and ***TT*** show very little interannual variability in sensitivity strength.

When we look at the time-normalized results, the story remains similar for evapotranspiration (although ***ETF*** becomes relatively even more dominant) but is somewhat different for streamflow, where the clear separation into groups of differing importance is no longer evident. We now see that snowmelt parameters ***CO*** and ***TT*** in the *Strong Importance* group are joined by the slow reservoir release rate coefficient ***K2*** and that relative ranking of these three parameters varies from year to year. Meanwhile, the soil parameters ***FRAC*** and ***FC*** form a group of *Moderate Importance*.

Overall, the results show that use of a strict *ranked order of importance* approach to characterizing sensitivity of model response to parameters is likely to be suspect and that the variability of results due to temporal variability in controlling factors such as system drivers should always be considered.

3.4. Time-Varying Parameter Importance Analysis: Treating Each Time Step Separately

Taking the issue of time-varying changes in parameter importance (due to changes in system drivers) to its logical limit, we now examine GSA results at the daily time scale (i.e., the integration time step of our model). Figure 8 shows the daily time variation of parameter importance indices for the Oldman for a representative 1-year period (October 2015 to September 2016). To indicate what is happening hydrometeorologically, the top subplot shows trajectories of observed precipitation, temperature, and streamflow. Each remaining subplot shows, for three selected parameters, the trajectory of the parameter importance index at each time step; for illustrative purposes the results are presented only for ***CO***, ***K2***, and ***ETF***.

3.4.1. Results Using Absolute Values of the Derivatives

Subplot (b) of Figure 8 shows, when no temporal normalization is applied, the trajectories of parameter importance indices that characterize the distributions of absolute values of the partial derivatives of streamflow and evapotranspiration with respect to the parameters. We see that streamflow is dominantly and strongly controlled by parameter ***CO*** (base temperature melt factor) throughout the year. In contrast, streamflow sensitivity to parameters ***K2*** (slow reservoir drainage rate) and ***ETF*** (evapotranspiration temperature-anomaly correction factor) is much smaller. As might be expected, given that it represents the base flow process, the parameter importance trajectory for parameter ***K2*** tends to be smoother and more slowly varying.

Subplot (c) shows corresponding results for evapotranspiration. It is again clear that sensitivity tends to be strongest to parameter ***ETF*** (evapotranspiration temperature-anomaly correction factor), with moderately strong sensitivity to ***CO*** (base temperature melt factor) as was also seen in section 3.3. Meanwhile, as should be expected, year-round sensitivity to parameter ***K2*** is zero. Finally, evapotranspiration becomes completely insensitive to all of the parameters during the winter season, when temperatures are low and the ground is frozen.

3.4.2. Results Using Temporally Normalized Absolute Derivatives

From subplots (d) and (e), which present the temporally normalized results, some interesting details become apparent. Because temporal normalization reduces the dominating effect of variations in strengths of the system input driving forces, the effect is to illuminate the *rank order* of relative importance rather than actual magnitude. Now the process importance at times when the system is not strongly driven (October to April) is more easily visible, with streamflow being strongly controlled by both ***K2*** and ***CO*** (followed by ***ETF***) between November and April, after which ***CO*** becomes clearly dominant. For evapotranspiration, the importance of ***ETF*** over ***CO*** becomes amplified.

3.4.3. Results Using Positively and Negatively Signed Subsets of the Derivatives

Figure 9 shows results when the distributions of positive and negative derivative values are treated separately, instead of lumped together. The left column shows results for Oldman, while the right column includes results for Banff. It is interesting that the sensitivity of streamflow to parameter ***CO*** (the dominant parameter) tends to be strongly positive in Oldman and strongly negative in Banff. Positive sensitivity indicates that increasing ***CO*** will tend to increase streamflow, while negative sensitivity indicates that increasing ***CO*** will tend to decrease streamflow. While intuition suggests that streamflow should be positively sensitive to ***CO***, we must also take into consideration the memory and dynamics of the system. With this in mind, a negative value for sensitivity implies simply that the available snow has been already melted and is therefore unable to contribute much to current streamflow.

For evapotranspiration, the dominant overall sensitivity of this flux to parameter ***ETF*** is fairly evenly balanced between positive and negative terms in both basins; compare this observation to Figure 8c, where the absolute value operation has been applied, lumping positive and negative derivative terms together.

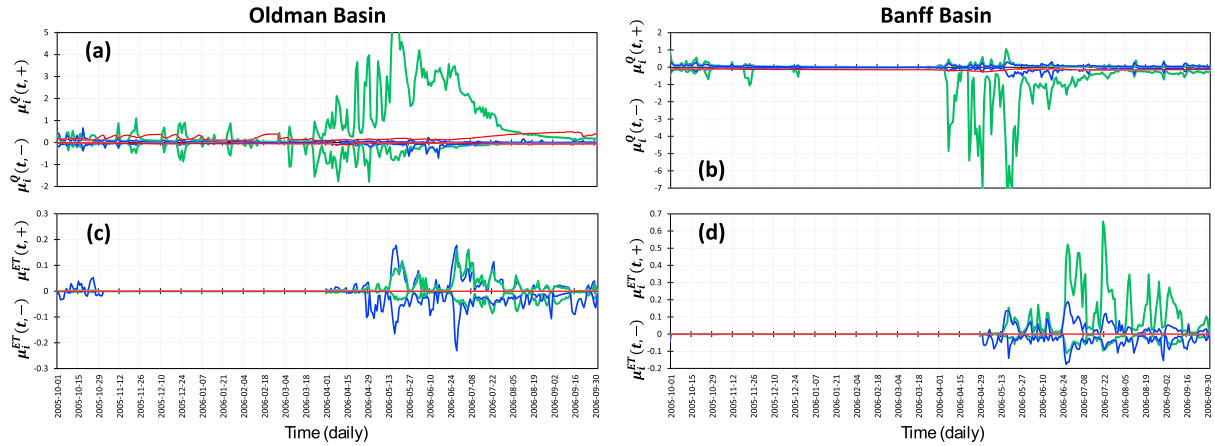

Figure 9. Plots showing GSM time-varying GSA for a representative period (October 2015 to September 2016) for the Oldman (left column) and Banff (right column) basins; (a and b) streamflow; (c and d) evapotranspiration. Here the parameter importance indices are computed separately for the *positive* and *negative* derivative values obtained for each time step. Only three selected parameters are shown (green = CO, blue = ETF, and red = K2).

3.5. Comparison With Performance Metric-Based Applications of *Sobol'* and *Morris*

Figure 10 compares results obtained using our GSM method to those obtained using metric-based (requiring availability of observed output data) applications of the *Morris* and *Sobol'* methods, for the Oldman. Here we plot the total period time-aggregate *parameter importance* indices obtained by our method on the *x* axes, against very accurate and stable estimates of the *Morris* and *Sobol'* parameter sensitivity indices on the *y* axes; the latter were generated as a by-product of the *VARS* methodology (Razavi and Gupta, 2016a, 2016b). This comparison cannot be performed for evapotranspiration (for which there is no observed data), highlighting a fundamental weakness of performance metric-based approaches. Results of our approach are shown for both regular and time-normalized indices.

The results indicate that *Morris* and *Sobol'* give similar time-aggregate parameter importance results (compare *y* coordinates in top and bottom rows). *Morris* ranks, in order of importance, the top three parameters as ***FRAC***, ***FC***, and ***CO*** (*Strong Importance*), followed by ***TT*** (*Moderate Importance*). *Sobol'* spaces out the relative importance of ***FRAC***, ***FC***, and ***CO*** (less closely grouped) and suggests that ***TT*** is only weakly important. Meanwhile, the other parameters appear together in a group of relatively low importance. Note that ***ETF***, ***LP***, and ***FC***, which exert strong controls on evapotranspiration, are not identified as being important by either method.

In contrast, the (metric-free) GSM approach ranks the top four parameters as being ***CO*** and ***TT*** (*Strong Importance*) and ***FRAC*** and ***FC*** (*Moderate Importance*). Note that time normalization also raises the parameter ***K2*** into the moderate-to-high importance category. There is general agreement with *Morris* and *Sobol'* about the remaining parameters being of low importance. Remember that our approach also identifies ***ETF*** (*Strong Importance*) and ***LP***, ***FC***, and ***CO*** (*Moderate Importance*) as exerting strong controls on evapotranspiration (see Figure 6).

Overall, even if we ignore the role of evapotranspiration, it is interesting that the metric-based applications of *Morris* and *Sobol'* give more importance to ***FRAC*** and ***FC*** that control soil processes, while our metric-free approach gives more importance to the parameters ***CO*** and ***TT*** that control snowmelt processes. Given the dominant snowmelt nature of the Oldman basin, the latter seems more intuitively correct.

3.6. Stability and Efficiency

It is important to assess the stability and efficiency of any GSA approach. In practice, the user needs to determine the base sample size $\boldsymbol{N_{Pts}}$ necessary to obtain robust results (see detailed discussion in Razavi and Gupta, 2016b). Figure 11 shows GSM total period time-aggregate importance indices computed using progressively larger base sample sizes in our HBV-SASK/Oldman example; because we applied a finite difference approximation to compute the partial derivatives at each location, this number should be multiplied by 11 (i.e., $\boldsymbol{N_\theta + 1}$) to obtain the total number of model runs. The results are highly stable, giving reliable results

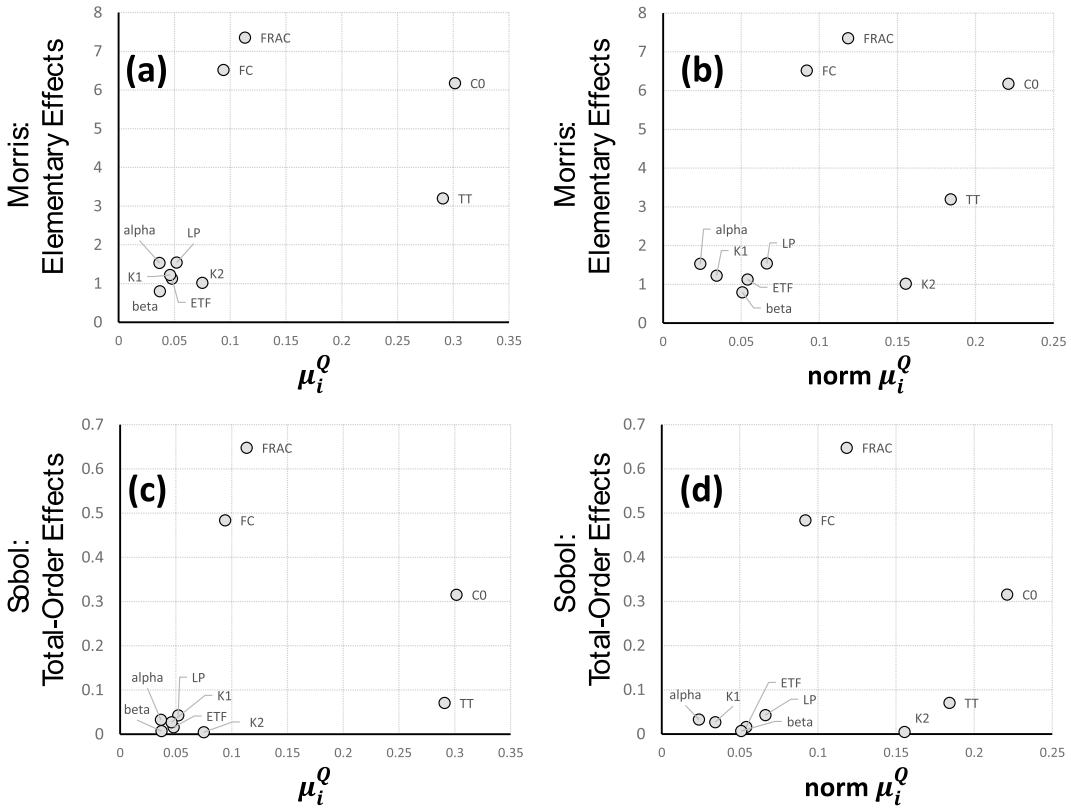

Figure 10. Plots comparing GSM total period time-aggregate SA results for streamflow to those obtained using performance-metric based *Morris* (top row) and *Sobol'* (bottom row) methods, for the Oldman basin. Results are shown for both raw derivative values (left column) and time-normalized values (right column).

(for both actual parameter importance index values and relative importance rankings) with $\boldsymbol{N_{Pts}}$ as low as 500–1,000 (i.e., 5,500–11,000 total model runs). Of the 10 parameters, only the results for ***K2*** for streamflow do not stabilize until about $\boldsymbol{N_{Pts}}$ = 600, and only the results of CO for evapotranspiration do not stabilize until about $\boldsymbol{N_{Pts}}$ = 1,000.

In contrast, Razavi and Gupta (2016b) have shown that typical implementations of the *Morris* and *Sobol'* approaches provide extremely poor reliabilities of parameter importance ranking for both low-dimensional (5-parameter) and relatively high-dimensional (45-parameter) models, even with relatively large sample sizes of 100,000 or more. We speculate that the GSM metric-free approach will generally be more efficient than metric-based methods because it does not lose important information via the aggregation and filtering operations performed by a performance metric. In other words, here we are extracting information directly from the model output partial derivatives $\boldsymbol{dY_t^k/d\theta_i}$ themselves (of which there are $\boldsymbol{N_{Pts}}$ * $\boldsymbol{T}$ values per parameter), rather than from the performance metric derivatives $\boldsymbol{dF^k/d\theta_i}$ (of which there are only $\boldsymbol{N_{Pts}}$ values per parameter).

4. Summary and Discussion

This paper revisits the problem of global factor importance analysis (i.e., commonly called global sensitivity analysis) for DESMs and proposes a reconceptualized basis for how such analyses should be performed. As DESMs become progressively more complex, it becomes increasingly important that meaningful information be obtainable regarding the importance of different factors controlling model behavior, while doing so at relatively low cost.

We argue that performance metric-based methods for GSA, when applied to parameters, should be recognized as actually being parameter identifiability analyses. Because the mathematical form of the metric

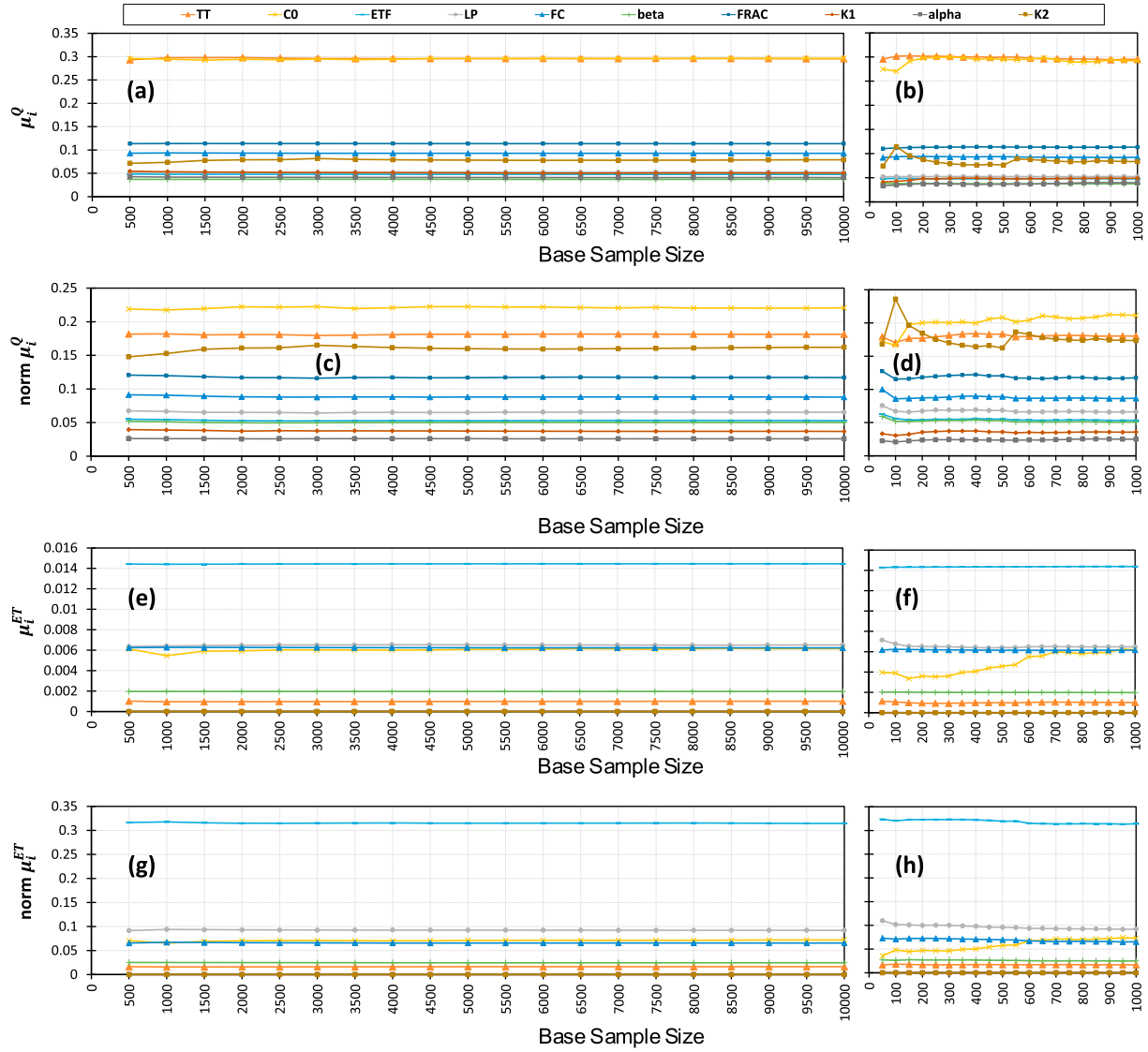

Figure 11. Plots illustrating very rapid convergence of the estimated values of GSM total period time-aggregate parameter importance indices with increasing base sample size; (a–d) Streamflow and (e–h) evapotranspiration. The subplots in the right column (b, d, f, and h) show detailed results for smaller base sample sizes. Note that this is a 10-parameter problem; to obtain total number of model runs, multiply by 11.

unavoidably distorts the information provided by the model about relative parameter importance, the relationship between identifiability and importance (i.e., sensitivity of outputs to parameter perturbations) is weakened. Further, because performance metric-based approaches depend on availability of system state/output data, the analysis obtained via such methods is necessarily incomplete and prone to uncertainty/errors in observations. Overall, it is a serious conceptual flaw to interpret the results of an identifiability analysis as being consistent and accurate indications of the sensitivity of model response to parameter perturbations.

We have therefore investigated the GSA problem from first principles, by starting from the theoretical basis for sensitivity, which is the magnitudes and signs of the partial derivatives of model output trajectories with respect to their controlling factors (here we have restricted attention to the model parameters). At a given parameter location, this information is provided by the two-dimensional local sensitivity matrix, (sometimes called the *Jacobian* matrix; Bard, 1976) that has $\boldsymbol{N_\theta}$ rows and $\boldsymbol{T}$ columns corresponding to the number of parameters and number of integration time steps (length of the simulation period), respectively. The analysis is made global by computing these values at a sufficiently well sampled representative set of $\boldsymbol{N_{Pts}}$ locations

uniformly distributed throughout the feasible parameter space, thereby obtaining a global sensitivity matrix from which statistical estimates of measures of relative parameter importance can be inferred. Because construction of the global sensitivity matrix does not depend on the availability of observed system output data, the approach enables a more comprehensive analysis of the properties of the model. Further, the approach can be readily used for retrospective-, predictive-, and scenario-type applications. The method can, of course, be extended to handle 1-D, 2-D, and 3-D spatial model responses by simply expanding the number of matrix dimensions.

In section 2.1.4, we discussed properties of the global sensitivity matrix that are expected to be typical for DESMs and proposed the use of parameter importance indices based on measures of central tendency (e.g., means) or dispersion (e.g., entropies) of the frequency distributions of absolute values of the partial derivatives; this proposal is based on our experimentally based observation that such distributions tend to be exponential-like and can be easily modified if some other statistical characterization seems more appropriate. However, for the case study reported here, we found the frequency distributions to indeed be exponential-like and accordingly used their means as summary indices of parameter importance. We also proposed that such indices be computed at various relevant time scales including (a) total period time-aggregate indices, (b) time series of annual time-aggregate indices, (c) dominance period time-aggregate indices, and (d) time series of integration time step indices.

Whereas the total period time-aggregate indices provide summary information regarding relative importance of various parameters, the time series of annual time-aggregate indices (and dominance period time-aggregate indices) reveal the inherent variability associated with such summary information, due to temporal variations in the strengths of the climatic factors controlling the system drivers. Meanwhile, the time series trajectories of parameter importance indices computed at the model integration time step can be very helpful for examining and understanding process level behavior of the model (a critically important role of GSA), and indices computed separately for positively and negatively signed derivatives can be especially helpful in this regard. For the Oldman, the time-varying analysis clearly shows the dominant effect of summer-winter seasonal variation on sensitivity of model response. It suggests that for regions with strongly seasonal climate, any time-aggregate GSA should be designed in a way to provide results separately for each season. Further, temporal grouping can be done for *dominance* periods such as melt/no-melt, rising-limb/falling-limb, or wet/dry (not shown here) to help tease out important information about model behavior. Temporal grouping can, in general, be different for different model outputs.

Of course, as with all other GSA methods, GSM results will unavoidably depend on the transformation and scaling chosen for the parameters and also possibly for the system outputs (particularly when different outputs or different study sites are to be compared). We recommend scaling parameters on to the range corresponds $[\mathbf{0}, \mathbf{1}]$ and use of log (or other) transformations for variables that typically vary over orders of magnitude.

Furthermore, we have discussed the potential value of time normalizing the results so as to draw out the relative importance of information provided at each model integration time step; for this, we proposed reweighting terms of the global sensitivity matrix to make the overall *total* sensitivity constant with time, so that each time step contributes equally to the analysis. The value of doing this was demonstrated to be particularly important when examining the time-varying dynamical nature of parameter importance, providing results that have properties similar to *ranking*.

In section 3, we demonstrated applicability of the GSM method by performing a parameter importance analysis for the HBV-SASK conceptual hydrologic model applied to the snow-dominated Oldman basin in Canada. A comparison of total period time-aggregate parameter importance indices generated by our metric-free approach against performance metric-based indices computed using the *Morris* and *Sobol'* methods showed that while all three approaches agreed regarding the *Low Importance* parameters, they differed regarding which parameters exert *strong* controls on model behavior. As explained in section 1.2, metric-based approaches provide biased results due to the way that they filter the sensitivity-relevant information—emphasizing the influence of information from regions containing *poor* parameters, while diminishing the influence of information from regions containing *good* parameters (i.e., that generate behaviors similar to those expressed by the observed data). This filtering role of metrics is amplified for parameters

having derivative distributions with larger absolute values, explaining why the metric-free and metric-based approaches tend to disagree regarding *moderate-to-high importance* parameters but tend to agree for low importance parameters.

Finally, it is important to point out that for any GSA method to provide scientifically useful insights regarding the dynamic behavioral functioning of the target environmental system, it is important to have already verified (before performing the GSA) that we have a reasonably good model—one that is structurally and functionally isomorphic to the real system. If the fidelity of the model is poor (see Clark et al., 2011), then the GSA will likely produce results that are artifacts and cannot be trusted. Of course, if the GSA provides results that run clearly counter to our intuitive understanding of the system, we may be able to use this information as a diagnostic tool to help us figure out what model improvements are necessary (Haghnegahdar et al., 2017). In our case, we did a crude preliminary analysis to characterize streamflow performance of the HBV-SASK model for the Oldman and Banff basins, by randomly sampling several thousand parameter locations across the feasible space and checking the frequency distributions of normalized MSEs so obtained (**NMSE = MSE/Var**($\boldsymbol{Q_{Obs}}$)). For both basins, about 10% of the obtained **NMSE** values were better (lower) than 0.5, with the best values approaching ~0.1.

5. Conclusions and Future Directions

This paper develops a theoretically sound Global Sensitivity Matrix method for time-aggregate and time-varying GSA, designed to assess relative importance of the parameters of DESMs. Illustrative testing has been limited to a preliminary proof-of-concept demonstration for a single model at two fairly similar study locations. Deeper understanding of the strengths and weaknesses of our method and of issues that may be encountered during practical application will require exhaustive testing on a variety of models at a large sample of locations (Gupta et al., 2014) and ongoing discussion among theoreticians and practitioners within the environmental systems modeling community. It will also be interesting to test and compare the time-varying parameter importance results against those provided by other published methods for time-varying sensitivity analysis.

We are also exploring logical extensions of the concepts developed here. In ongoing work, we are generalizing the GSM method to accommodate alternative choices of ***R*** (including targeted or compressed characterizations of the model response and even model performance metrics) and/or the alternative basic ways of computing the sensitivity coefficients (finite differences, derivatives, and discrete conditions). The Generalized GSM (GGSM) approach will therefore enable computation of analogous time-aggregate and time-varying factor importance indices using alternative analytical techniques, including ones that are not *derivative-based* (e.g., *Sobol'*, *VARS, PAWN*, and *moment-based*). GGSM is being programmed into a comprehensive VARS-TOOL software package (Razavi et al., 2018) wherein the STAR sampling strategy (Razavi & Gupta, 2016b) provides an efficient basis for generating the required sample locations. Accordingly, a single STAR-VARS experiment, where the star centers are generated using progressive Latin hypercube sampling (Sheikholeslami & Razavi, 2017), can be used to simultaneously generate a variety of methodological estimates of factor importance (e.g., *Morris, Sobol',* and the more recent *IVARS* metrics), along with the estimates developed in this paper. The VARS-TOOL package can be obtained from the second author by request; a website for its dissemination has been constructed (http://vars-tool.com).

We believe that future work should be devoted to developing a logical framework that helps the practitioner to properly select among the various choices involved in implementation of GSA for their specific application and/or decision context. These choices include (a) appropriate choice of ***R***, (b) appropriate transformations of ***R*** and ***C***, (c) whether/when to focus on the *responses* (as in *Sobol', FAST*, *PAWN,* and *moment-based*), or *derivatives* of the responses (as in *Morris'*, *DELSA,* and *GSM*), or both (as in *VARS*), and (d) which properties of the frequency distributions obtained by global sampling are relevant (e.g., means, variances, higher-order moments, and shapes of the cumulative distribution functions). As mentioned in section 1.2, one promising area is the use of diagnostically informative signature properties of system response to (potentially) strengthen the role of SA in evaluation of alternative model structural hypotheses. Another is the need for strategies that overcome the curse of dimensionality (and associated computational costs) thereby enabling GSA to be applied to very high dimensional DESMs; an interesting *grouping* solution to this problem was recently proposed by Sheikholeslami et al. (2018).

As always, we invite discussion and collaboration with others interested in these and related issues of system identification and model development. In particular, we seek collaborations to test and compare a variety of GSA methods for large samples of environmental system types, model types, and study locations across a variety of climatic/environmental regimes, so that our approach to assessing parameter importance can be further tested, improved, and refined.

Acknowledgments
The first author received partial support from the *Australian Research Council* through the Centre of Excellence for Climate System Science (grant CE110001028). The second author was supported in part by a NSERC (Natural Sciences and Engineering Research Council of Canada) Discovery Grant. The data, model, and VARS toolbox used to support this paper are available for download at http://vars-tool.com or from the second author (saman.razavi@usask.ca) on request.

References

Abily, M., Bertrand, N., Delestre, O., Gourbesville, P., & Duluc, C. M. (2016). Spatial global sensitivity analysis of high resolution classified topographic data use in 2D urban flood modeling. *Environmental Modeling& Software*, *77*, 183–195.

Bard, Y. (1976). *Nonlinear parameter estimation*. New York and London: Academic Press.

Borgonovo, E., Lu, X., Plischke, 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*, 7933–7950. https://doi.org/10.1002/2017WR020767

Box, G. E. P., & Cox., D. (1974). An analysis of transformations. *Journal of the Royal Statistical Society*, *26*, 211–252.

Boyle, D. P., Gupta, H. V., & Sorooshian, S. (2000). Towards improved calibration of hydrologic models: Combining the strengths of manual and automatic methods. *Water Resources Research*, *36*(12), 3663–3674. https://doi.org/10.1029/2000WR900207

Campbell, K., McKay, M. D., & Williams, B. J. (2006). Sensitivity analysis when model outputs are functions. *Reliability Engineering & System Safety*, *91*(10–11), 1468–1472.

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

Cibin, R., Sudheer, K. P., & Chaubey, I. (2010). Sensitivity and identifiability of stream flow generation parameters of the SWAT model. *Hydrological Processes*, *24*, 1133–1148.

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

Cloke, H., Pappenberger, F., & Renaud, J. P. (2008). Multi-method global sensitivity analysis (MMGSA) for modeling floodplain hydrological processes. *Hydrological Processes: An International Journal*, *22*(11), 1660–1674.

Cukier, R., Levine, H., & Shuler, K. (1978). Nonlinear sensitivity analysis of multiparameter model systems. *Journal of Computational Physics*, *26*, 1–42.

Dell'Oca, A., Riva, M., & Guadagnini, A. (2017). Moment-based metrics for global sensitivity analysis of hydrological systems. *Hydrology and Earth System Sciences Discussions*. https://doi.org/10.5194/hess-2017-90-ACI

Demaria, E. M., Nijssen, B., & Wagener, T. (2007). Monte Carlo sensitivity analysis of land surface parameters using the Variable Infiltration Capacity model. *Journal of Geophysical Research*, *112*, D11113. https://doi.org/10.1029/2006JD007534

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez-Baquero, G. F. (2009). Decomposition of the mean squared error & NSE performance criteria: Implications for improving hydrological modeling. *Journal of Hydrology*, *377*, 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003

Gupta, H. V., & Nearing, G. S. (2014). Debates—The future of hydrological sciences: A (common) path forward? Using models and data to learn: A systems theoretic perspective on the future of hydrological science, Invited Commentary. *Water Resources Research*, *50*, 5351–5359. https://doi.org/10.1002/2013WR015096

Gupta, H. V., Perrin, C., Blöschl, G., Montanari, A., Kumar, R., Clark, M., & Andréassian, V. (2014). Large-sample hydrology: A need to balance depth with breadth. *Hydrology and Earth Systems Science*, *18*, 463–477. https://doi.org/10.5194/hess-18-463-2014

Gupta, H. V., Wagener, T., & Liu, Y. Q. (2008). Reconciling theory with observations: Towards a diagnostic approach to model evaluation. *Hydrological Processes*, *22*(18), 3802–3813. https://doi.org/10.1002/hyp.6989

Gupta, V. K., & Sorooshian, S. (1985). The automatic calibration of conceptual catchment models using derivative-based optimization algorithms. *Water Resources Research*, *21*(4), 473–486. https://doi.org/10.1029/WR021i004p00473

Guse, B., Pfannerstill, M., Strauch, M., Reusser, D., Lüdtke, S., Volk, M., et al. (2016). On characterizing the temporal dominance patterns of model parameters and processes. *Hydrological Processes*, *30*, 2255–2270. https://doi.org/10.1002/hyp.10764

Guse, B., Reusser, D. E., & Fohrer, N. (2014). How to improve the representation of hydrological processes in SWAT for a lowland catchment—Temporal analysis of parameter sensitivity and model performance. *Hydrological Processes*, *28*, 2651–2670. https://doi.org/10.1002/hyp.977

Haghnegahdar, A., & Razavi, S. (2017). Insights into sensitivity analysis of earth and environmental systems models: On the impact of parameter perturbation scale. *Environmental Modeling & Software*, *95*, 115–131.

Haghnegahdar, A., Razavi, S., Yassin, F., & Wheater, H. (2017). Multi-criteria sensitivity analysis as a diagnostic tool for understanding model behavior and characterizing model uncertainty. *Hydrological Processes*. https://doi.org/10.1002/hyp.11358

Herman, J. D., Kollat, J. B., Reed, P. M., & Wagener, T. (2013). From maps to movies: High resolution time-varying sensitivity analysis for spatially distributed watershed models. *Hydrology and Earth System Sciences*, *17*, 5109–5125.

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*, 1400–1414. https://doi.org/10.1002/wrcr.20124

Lamboni, M., Makowski, D., Lehuger, S., Gabrielle, B., & Monod, H. (2009). Multivariate global sensitivity analysis for dynamic crop models. *Field Crops Research*, *113*(3), 312–320.

Lamboni, M., Monod, H., & Makowski, D. (2011). Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models. *Reliability Engineering & System Safety*, *96*(4), 450–459.

Le Cozannet, G., Rohmer, J., Cazenave, A., Idier, D., van de Wal, R., de Winter, R., et al. (2015). Evaluating uncertainties of future marine flooding occurrence as sea-level rises. *Environmental Modeling& Software*, *73*, 44–56.

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

Marrel, A., Iooss, B., Jullien, M., Laurent, B., & Volkova, E. (2011). Global sensitivity analysis for models with spatially dependent outputs. *Environmetrics*, *22*(3), 383–397.

Massmann, C., Wagener, T., & Holzmann, H. (2014). A new approach to visualizing time-varying sensitivity indices for environmental model diagnostics across evaluation time-scales. *Environmental Modeling and Software*, *51*, 190–194. https://doi.org/10.1016/j.envsoft.2013.09.033

Moench, A. (1994). Specific yield as determined by type-curve analysis of aquifer-test data. *Ground Water*, *32*, 949–957.

Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics*, *33*(2), 161e174.
Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models. Part I: A discussion of principles. *Journal of Hydrology*, *10*, 282–290.
Pappenberger, F., Beven, K. J., Ratto, M., & Matgen, P. (2008). Multi-method global sensitivity analysis of flood inundation models. *Advances in Water Resources*, *31*(1), 1–14.
Pianosi, F., Beven, K., Freer, J., Hall, J. W., Rougier, J., Stephenson, D. B., & Wagener, T. (2016). Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environmental Modeling& Software*, *79*, 214–232.
Pianosi, F., & Wagener, T. (2015). A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. *Environmental Modeling & Software*, *67*, 1–11. https://doi.org/10.1016/j.envsoft.2015.01.004
Pianosi, F., & Wagener, T. (2016). Understanding the time-varying importance of different uncertainty sources in hydrological modeling using global sensitivity analysis. *Hydrological Processes*, *30*, 3991–4003. https://doi.org/10.1002/hyp.10968
Rakovec, O., Hill, M. C., Clark, M. P., Weerts, A. H., Teuling, A. J., & Uijlenhoet, R. (2014). Distributed evaluation of local sensitivity analysis (DELSA), with application to hydrologic models. *Water Resources Research*, *50*, 409–426. https://doi.org/10.1002/2013WR014063
Razavi, S., & Gupta, H. V. (2015). What do we mean by sensitivity analysis? The need for a comprehensive characterization of "global" sensitivity in Earth and environmental systems models. *Water Resources Research*, *51*, 3070–3092. https://doi.org/10.1002/2014WR016527
Razavi, S., & Gupta, H. V. (2016a). A new framework for comprehensive, robust, and efficient global sensitivity analysis: Part I—Theory. *Water Resources Research*, *52*, 423–439. https://doi.org/10.1002/2015WR017558
Razavi, S., & Gupta, H. V. (2016b). A new framework for comprehensive, robust, and efficient global sensitivity analysis: Part II—Applications. *Water Resources Research*, *52*, 440–455. https://doi.org/10.1002/2015WR017559
Razavi, S., Sheikholeslami, R., Gupta, H., & Haghnegahdar, A. (2018). VARS-TOOL: A toolbox for comprehensive, efficient, and robust global sensitivity analysis. *Environmental Modelling and Software*, https://doi.org/10.1016/j.envsoft.2018.10.005
Reusser, D. E., Buytaert, W., & Zehe, E. (2011). Temporal dynamics of model parameter sensitivity for computationally expensive models with FAST (Fourier Amplitude Sensitivity Test). *Water Resources Research*, *47*, W07551. https://doi.org/10.1029/2010WR009947
Reusser, D. E., & Zehe, E. (2011). Inferring model structural deficits by analyzing temporal dynamics of model performance and parameter sensitivity. *Water Resources Research*, *47*, W07550. https://doi.org/10.1029/2010WR009946
Rosolem, R., Gupta, H. V., Shuttleworth, W. J., Zeng, X., & de Goncalves, L. G. G. (2012). A fully multiple-criteria implementation of the Sobol method for parameter sensitivity analysis. *Journal of Geophysical Research*, *117*, D07103. https://doi.org/10.1029/2011JD016355
Rupp, D. E., & Selker, J. S. (2006). Information, artifacts, and noise in dQ/dt—Q recession analysis. *Advances in Water Resources*, *29*(2), 154–160.
Saltelli A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., et al. (2008). *Global sensitivity analysis: The primer*. Hoboken, NJ: John Wiley.
Savage, J. T. S., Pianosi, F., Bates, P., Freer, J., & Wagener, T. (2016). Quantifying the importance of spatial resolution and other factors through global sensitivity analysis of a flood inundation model. *Water Resources Research*, *52*, 9146–9163. https://doi.org/10.1002/2015WR018198
Sheikholeslami, R., & Razavi, S. (2017). Progressive Latin Hypercube Sampling: An efficient approach for robust sampling-based analysis of environmental models. *Environmental Modeling & Software*, *93*, 109–126.
Sheikholeslami, R., Razavi, S., Gupta, H., 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. *Environmental Modeling & Software*. https://doi.org/10.1016/j.envsoft.2018.09.002, in press
Shin, M.-J., Guillaume, J. H., Croke, B. F., & Jakeman, A. J. (2013). Addressing ten questions about conceptual rainfall–runoff models with global sensitivity analyses in R. *Journal of Hydrology*, *503*, 135–152.
Sieber, A., & Uhlenbrook, S. (2005). Sensitivity analyses of a distributed catchment model to verify the model structure. *Journal of Hydrology*, *310*(1–4), 216–235.
Sivapalan, M., Blöschl, G., Zhang, L., & Vertessy, R. (2003). Downward approach to hydrological prediction. *Hydrological Processes*, *17*, 2101–2111. https://doi.org/10.1002/hyp.1425
Sobol', I., & Levitan, Y. L. (1999). On the use of variance reducing multipliers in Monte Carlo computations of a global sensitivity index. *Computer Physics Communications*, *117*(1), 52–61.
Sobol', I. M. (1990). On sensitivity estimation for nonlinear mathematical models. *Matematicheskoe Modelirovanie*, *2*(1), 112–118.
Sobol', I. M., & Kucherenko, S. (2009). Derivative based global sensitivity measures and their link with global sensitivity indices. *Mathematics and Computers in Simulation*, *79*(10), 3009–3017.
Van Griensven, A., Meixner, T., Grunwald, S., Bishop, T., Diluzio, M., & Srinivasan, R. (2006). A global sensitivity analysis tool for the parameters of multi-variable catchment models. *Journal of Hydrology*, *324*(1–4), 10–23.
Vogel, R. M., & Sankarasubramanian, A. (2003). The validation of a watershed model without calibration. *Water Resources Research*, *39*(10), 1292. https://doi.org/10.1029/2002WR001940
Vrugt, J. A., Gupta, H. V., Bastidas, L. A., Bouten, W., & Sorooshian, S. (2003). Effective and efficient algorithm for multi-objective optimization of hydrologic models. *Water Resources Research*, *39*(8), 1214. https://doi.org/10.1029/2002WR001746
Van Werkhoven, K., Wagener, T., Reed, P., & Tang, Y. (2008a). Characterization of watershed model behavior across a hydroclimatic gradient. *Water Resources Research*, *44*, W01429. https://doi.org/10.1029/2007WR006271
Van Werkhoven, K., Wagener, T., Reed, P., & Tang, Y. (2008b). Rainfall characteristics define the value of streamflow observations for distributed watershed model identification. *Geophysical Research Letters*, *35*, L11403. https://doi.org/10.1029/2008GL034162
Wagener, T., McIntyre, N., Less, M. J., Wheater, H. S., & Gupta, H. V. (2003). Towards reduced uncertainty in conceptual rainfall-runoff modeling: Dynamic identifiability analysis. *Hydrological Processes*, *17*(2), 455–476. https://doi.org/10.1002/hyp.1135
Yilmaz, K. K., Gupta, H. V., & Wagener, T. (2008). A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model. *Water Resources Research*, *44*, W09417. https://doi.org/10.1029/2007WR006716

Revisiting the Basis of Sensitivity Analysis for Dynamical Earth System Models

Gupta, Hoshin V.; Razavi, Saman

01 mingxi zhang Page 1
12/1/2024 6:27

02 mingxi zhang Page 1
12/1/2024 6:29

03 mingxi zhang Page 1
12/1/2024 6:29

04 mingxi zhang Page 1
15/1/2024 3:35

05 mingxi zhang Page 7
15/1/2024 3:37