Distributed evaluation of local sensitivity analysis (DELSA), with application to hydrologic models

O. Rakovec,[1,2] M. C. Hill,[3] M. P. Clark,[4] A. H. Weerts,[1,5] A. J. Teuling,[1] and R. Uijlenhoet[1]

Received 3 May 2013; revised 18 October 2013; accepted 13 December 2013.

[1] This paper presents a hybrid local-global sensitivity analysis method termed the Distributed Evaluation of Local Sensitivity Analysis (DELSA), which is used here to identify important and unimportant parameters and evaluate how model parameter importance changes as parameter values change. DELSA uses derivative-based "local" methods to obtain the distribution of parameter sensitivity across the parameter space, which promotes consideration of sensitivity analysis results in the context of simulated dynamics. This work presents DELSA, discusses how it relates to existing methods, and uses two hydrologic test cases to compare its performance with the popular global, variance-based Sobol' method. The first test case is a simple nonlinear reservoir model with two parameters. The second test case involves five alternative "bucket-style" hydrologic models with up to 14 parameters applied to a medium-sized catchment (200 km^2) in the Belgian Ardennes. Results show that in both examples, Sobol' and DELSA identify similar important and unimportant parameters, with DELSA enabling more detailed insight at much lower computational cost. For example, in the real-world problem the time delay in runoff is the most important parameter in all models, but DELSA shows that for about 20% of parameter sets it is not important at all and alternative mechanisms and parameters dominate. Moreover, the time delay was identified as important in regions producing poor model fits, whereas other parameters were identified as more important in regions of the parameter space producing better model fits. The ability to understand how parameter importance varies through parameter space is critical to inform decisions about, for example, additional data collection and model development. The ability to perform such analyses with modest computational requirements provides exciting opportunities to evaluate complicated models as well as many alternative models.

Citation: Rakovec, O., M. C. Hill, M. P. Clark, A. H. Weerts, A. J. Teuling, and R. Uijlenhoet (2014), Distributed evaluation of local sensitivity analysis (DELSA), with application to hydrologic models, *Water Resour. Res.*, *50*, doi:10.1002/2013WR014063.

1. Introduction

[2] The primary aim of sensitivity analysis is to identify how different components of a model affect model output. Results can affect model calibration, uncertainty evaluation, and risk assessment [e.g., *Katz*, 1999; *Oakley and O'Hagan*, 2004; *Saltelli et al*., 2008; *Kucherenko et al*., 2009; *Ohler et al.*, 2013; *Plischke et al*., 2013]. Sensitivity analyses can be conducted to evaluate model structure and forcing [e.g., *Doherty and Welter*, 2010; *Gupta et al*., 2012; *Foglia et al*., 2013], yet most commonly, and in this work, the analysis is focused on evaluating how model parameters (also called factors) affect model output. Focusing on parameter sensitivity analysis is less limiting than it might seem because parameters can be defined to control many aspects of a model, and thus parameter sensitivity analysis methods can be used to understand the role played by, for example, model structure and forcing. Once parameters are defined, parameter sensitivity analysis is used to identify parameters important and unimportant to simulated processes, inverse modeling metrics, predictions, and quantified uncertainty and risk [e.g., *Van Werkhoven et al*., 2008; *Saltelli et al*., 2008; *Plischke et al*., 2013]. Within these broad goals, parameter sensitivity analysis can be used to (a) detect when increasing model complexity can no longer be supported by observations and whether it is likely to affect predictions of interest [e.g., *Saltelli et al*., 1999; *Van Werkhoven et al*., 2008; *Rosolem et al*., 2012; *Gupta et al*., 2012], (b) reduce the time of model calibration by focusing estimation efforts

[1]Hydrology and Quantitative Water Management Group, Department of Environmental Sciences, Wageningen University, Wageningen, Netherlands.
[2]Now at UFZ - Helmholtz Centre for Environmental Research, Leipzig, Germany.
[3]U.S. Geological Survey—Water Resources Division, Boulder, Colorado, USA.
[4]Research Applications Laboratory, National Center for Atmospheric Research, Boulder, Colorado, USA.
[5]Deltares, Delft, Netherlands.

Corresponding author: O. Rakovec, UFZ - Helmholtz Centre for Environmental Research, Leipzig 04318, Germany. (oldrich.rakovec@ufz.de)

0043-1397/14/10.1002/2013WR014063

on parameters important to calibration metrics and predictions [e.g., *Anderman et al*., 1996; *Hamm et al*., 2006; *Zambrano-Bigiarini and Rojas*, 2013], (c) determine priorities for theoretical and site-specific model development [e.g., *Hill and Tiedeman*, 2007; *Saltelli et al*., 2008; *Kavetski and Clark*, 2010], and (d) identify advantageous placement and timing of new measurements [e.g., *Musters and Bouten*, 2000; *Weerts et al*., 2001; *Vrugt et al*., 2001; *Tiedeman et al*., 2003, 2004; *Tonkin et al*., 2007; *Fienen et al*., 2010]. Sensitivity analysis evaluated for different periods has been considered by, e.g., *Wagener et al*. [2003], *Cloke et al*. [2007], and *Herman et al*. [2013a].

[3] The considerable potential utility of sensitivity analysis methods is sometimes difficult to attain in practice because many sensitivity analysis methods currently popular in environmental modeling require considerable computational effort. To discuss this issue, we use the common classification of sensitivity analysis methods: global and local.

[4] Global methods calculate sensitivity measures using parameter samples obtained from a defined parameter space. For each parameter sample, the model is run one or more times to obtain simulated results. Global methods provide stable results because they produce measures of parameter importance that are averaged over the range of the parameter space from which samples are obtained. However, of interest is that different global sensitivity analysis methods can lead to completely different importance rankings, as shown, for example, by *Pappenberger et al*. [2008].

[5] When considered from the perspective of uncertainty analysis, global methods can be moment independent [e.g., *Park and Ahn*, 1994; *Chun et al*., 2000; *Borgonovo*, 2007] or based on decomposition of the variance of the model output into variance contributions from individual parameters and parameter combinations [*Saltelli et al*., 2008]. The latter are considered in this work. Many global methods account for the effect of nonlinearity, including the effect of parameter interactions of predictions [e.g., *Box and Tiao*, 1992; *Saltelli*, 2002; *Van Werkhoven et al*., 2008; *Rosolem et al*., 2012].

[6] A popular variance-based global method is the Sobol' method [e.g., *Sobol'*, 2001; *Saltelli*, 2002], which has been applied in many environmental studies [e.g., *Hamm et al*., 2006; *Pappenberger et al*., 2006; *Cloke et al*., 2007; *Tang et al*., 2007; *Pappenberger et al*., 2008; *Van Werkhoven et al*., 2008; *Saltelli et al*., 2008; *Kavetski and Clark*, 2010; *Nossent et al*., 2011; *Massmann and Holzmann*, 2012; *Herman et al*., 2013a, 2013b] and is often used in method comparisons [e.g., *Borgonovo*, 2006; *Cloke et al*., 2007; *Pappenberger et al*., 2008]. Thus, Sobol' can be used as a standard for comparison that allows alternative methods to be indirectly compared to each other. Many global methods, and especially the Sobol' method, require a large number of model runs (often 10,000 or more) and thus can be computationally demanding or prohibitive.

[7] Alternative global methods such as Fourier amplitude sensitivity testing (FAST) [*Cukier et al*., 1973, 1975, 1978; *Saltelli et al*., 1999], the method of Morris (MoM) [*Morris*, 1991; *Zhan et al*., 2013; *Herman et al*., 2013b], regional sensitivity analysis (RSA) [*Hornberger and Spear*, 1981; *Freer et al*., 1996], and the delta method [*Borgonovo*, 2007] provide more frugal alternatives, but often remain a computational challenge: for example, FAST and the delta method require at least one solution at each sampling point and convergence properties are such that the number of sampling points is large; MoM is often conducted with multiple parameter increments for each parameter at each sampling point.

[8] Local sensitivity analysis methods are typically based on the gradients (derivatives) of the model output with respect to parameter values evaluated at a single location in the parameter space [e.g., *Hill and Tiedeman*, 2007; *Oliver et al*., 2008]. Parameter interactions on predictions can be accounted for using, for example, the method described by *Sobol' and Kucherenko* [2010]. The convenience of local methods has resulted in their considerable use [e.g., *D'Agnese et al*., 1999; *Kunstmann et al*., 2002]. However, their applicability to nonlinear models, including models with spurious results, is of concern [*Saltelli et al*., 2008] because single-point application of local methods can identify dramatically different important and unimportant parameters in different parts of the parameter space, and results could mislead modelers and users of model results. Comparisons of global and local methods, including *Helton* [1993], *Tang et al*. [2007], *Foglia et al*. [2007], *Kucherenko et al*. [2009], *Sobol' and Kucherenko* [2009, 2010], *Delenne et al*. [2012], and *Li et al*. [2013] provide mixed results. For example, *Foglia et al*. [2007] used a hydrologic model and the nonlocal cross-validation method to show that local methods provided useful information on the sensitivities of 35 parameters. *Tang et al*. [2007] compared multiple sensitivity analysis methods applied to the lumped Sacramento soil moisture accounting model and found that local methods provided very different results from the global methods—however, the local methods were not scaled to be dimensionless, which potentially explains the different model sensitivities. *Delenne et al*. [2012] showed that despite the nonlinearity of river flow processes simulated with a one-dimensional hydrodynamic model, a local approach yielded similar results to a global approach, even in cases of large parameter uncertainty. *Hamm et al*. [2006] compared global and local approaches to evaluate the effect of parameter values on the probability of hydrologically induced slope stability. They highlighted the robustness of the Sobol' method, and demonstrated that it accounted for parameter interactions neglected by the local method they used. To the extent that such irregularities are uncharacteristic of the system involved, code changes that reduce the solution irregularities may be advantageous. To the extent that they are realistic or numerically unavoidable, they present a serious problem for the use of local sensitivity analysis methods.

[9] An unresolved problem in sensitivity analysis is whether the averaging over parameter space inherent in global methods can obscure important aspects of parameter sensitivity. If sensitivity analysis results vary across parameter space, the robustness of decisions and measures based on sensitivity analysis becomes an issue. For example, data collection strategy decisions and uncertainty quantification may change significantly if different parameter ranges were used in a global analysis. Considerable variability would suggest the importance of identifying the most relevant

sensitivity analysis results. In some cases more relevant results might be thought to be associated with models that fit measurements better.

[10] To reveal how sensitivity analysis measures vary as parameter values change, we present and evaluate the hybrid local-global Distributed Evaluation of Local Sensitivity Analysis (DELSA) method, in which multiple evaluations of local parameter sensitivity are distributed throughout parameter space. DELSA integrates methodological features from three existing sensitivity analysis methods: the Method of Morris (MoM) [*Morris*, 1991; *Zhan et al*., 2013; *Herman et al*., 2013b], the Sobol' method [*Sobol'*, 2001; *Van Werkhoven et al*., 2008], and regional sensitivity analysis (RSA) [*Hornberger and Spear*, 1981; *Freer et al*., 1996]. We are not the first to consider the distribution of sensitivities across parameter space; *Saltelli et al*. [2008] call it "factor mapping," but expressly exclude derivative-based "local" methods as a useful tool. Our work suggests the potential utility of local methods in the context of this "multiscale" evaluation of parameter sensitivity.

[11] Our evaluation includes comparisons with the global variance-based Sobol' method using first-order sensitivity analysis, though DELSA is broadly applicable. To enable comparison, the Sobol' method is sometimes applied at a smaller than normal scale; and sometimes local DELSA measures are averaged over large-scale parameter space. Computational costs are also compared.

[12] We begin by briefly presenting the Sobol' method, before introducing DELSA. We compare the methods using two examples. The first example is a simple two-parameter reservoir model. The second example involves five alternative bucket-style hydrologic models applied to the Lasnenville catchment in the Belgian Ardennes (Western Europe).

2. Sensitivity Analysis Methods

2.1. Global Variance-Based Approach

[13] The global Sobol' method [*Sobol'*, 2001; *Saltelli*, 2002; *Sobol' and Kucherenko*, 2010] decomposes the variance of a metric describing model output (in this work, mean simulated reservoir storage or root-mean-square-error between the simulated and observed streamflow) into contributions from individual parameters. Briefly, consider a model f and a vector $\boldsymbol{\theta}$ of k model parameters, which yield a metric Ψ describing model output:

$$\Psi = f(\boldsymbol{\theta}) = f(\theta_1, \theta_2, \ldots, \theta_k), \tag{1}$$

where the range of each parameter θ_j is defined as $\theta_{j,min}$ to $\theta_{j,max}$. Equation (1) can be decomposed as described by *Sobol' and Kucherenko* [2010] to obtain the general variance decomposition scheme

$$V(\Psi) = \sum_j V_j + \sum_j \sum_{m>j} V_{jm} + \cdots + V_{1,2,\ldots,k}, \tag{2}$$

where $V(\Psi)$ is the total prior (unconditional) variance of Ψ, V_j are the first-order terms, V_{jm} are the second-order terms, etc. The Sobol' sensitivity indices are calculated as the ratio between partial variances and the total variance and are conveniently scaled within [0,1]. In this work, we use the first-order Sobol' sensitivity index (S_1^j) calculated as

$$S_1^j = \frac{V_j}{V(\Psi)}. \tag{3}$$

[14] It represents the main effect of a parameter θ_j on the total variance $V(\Psi)$ and considers no parameter interactions relative to the predictions. Following *Homma and Saltelli* [1996] and *Saltelli* [2002], the first-order variance term V_j can be estimated using the following equations:

$$V_j = \hat{U}_j - \hat{E}^2(\Psi), \tag{4}$$

$$\hat{U}_j = \frac{1}{N_S - 1} \sum_{r=1}^{N_S} (\Psi_r \times \Psi_r^j), \tag{5}$$

$$\hat{E}^2(\Psi) = \left(\frac{1}{N_S} \sum_{r=1}^{N_S} \Psi_r \right)^2, \tag{6}$$

where $\hat{U}_j$ is the "perturbed variance," $\hat{E}^2(\Psi)$ is the squared mean, and N_S is the number of parameter samples used in the Sobol' procedure, which should not be confused with the number of model parameters, k. Ψ_r and Ψ_r^j are the model performance indices of the rth parameter set:

$$\Psi_r = f(\theta_{r,1}, \theta_{r,2}, \ldots, \theta_{r,k}), \tag{7}$$

$$\Psi_r^j = f(\theta'_{r,1}, \theta'_{r,2}, \ldots, \theta'_{r,j-1}, \theta_{r,j}, \theta'_{r,j+1}, \ldots, \theta'_{r,k}). \tag{8}$$

[15] Ψ_r^j in equation (8) is constructed from equation (7) in such a way that all parameters except the jth parameter are perturbed by inserting the value of the parameters generated in another randomly selected Sobol' set of parameters. Finally, the total variance $V(\Psi)$ is estimated by $\hat{V}(\Psi)$:

$$\hat{V}(\Psi) = \hat{E}(\Psi^2) - \hat{E}^2(\Psi) = \frac{1}{N_S} \sum_{r=1}^{N_S} \Psi_r^2 - \left(\frac{1}{N_S} \sum_{r=1}^{N_S} \Psi_r \right)^2. \tag{9}$$

[16] The calculation cost for the first-order Sobol' index, in terms of the number of model runs, is N_S $(k + 1)$. The ranges are the only parameter information used. The parameters are sampled using the quasi-random Sobol' sequence, which draws realizations from the least visited locations relative to previous samples in the parameter space [*Bratley and Fox*, 1988] and represents an alternative to Latin hypercube sampling.

[17] If Ψ is a purely additive function of the parameters, so that each additive term is a function of at most one parameter, the sum of the first-order Sobol' sensitivity indices equals one $\left(\sum_{j=1}^{k} S_1^j = 1 \right)$. The sampling uncertainty of S_1^j is evaluated by bootstrapping with resampling [e.g., *Efron and Tibshirani*, 1993; *Archer et al*., 1997]. In this study, the number of bootstrap samples is 1000.

2.2. Distributed Evaluation of Local Sensitivity Analysis (DELSA)

[18] The local sensitivity analysis application, upon which DELSA expands, is based on the sensitivity (gradient) of model output or performance index Ψ with respect to the jth parameter value

$$\left.\frac{\partial \Psi}{\partial \theta_j}\right|_l, \tag{10}$$

where l defines the set of parameter values at which the derivative is calculated. Equation (10) is estimated using a forward difference approach using 1% change in the parameter value $[\Psi(\theta_j+0.01\theta_j)-\Psi(\theta_j)]/[0.01\theta_j]$. The choice of the 1% parameter change was also compared with 0.1% and 10% parameter change, but had a marginal effect (not shown).

[19] Here we use the DELSA concept with a local sensitivity statistic to provide a clear comparison with Sobol'.

[20] The local equation for total variance comparable to equations (2) and (9) is the prediction variance $V_L(\Psi)$ (*Seber and Wild* [1989, p. 191]; *Draper and Smith* [1998, p. 129–130]; *Aster et al*. [2013, p. 224]; *Lu et al*. [2012] discuss the relation between frequentist and Bayesian calculation of variance). $V_L(\Psi)$ is calculated as

$$V_L(\Psi)=\left(\left.\frac{\partial \Psi}{\partial \boldsymbol{\theta}}\right|_l\right)^T(\boldsymbol{X}^T\boldsymbol{\omega}\boldsymbol{X})^{-1}\left(\left.\frac{\partial \Psi}{\partial \boldsymbol{\theta}}\right|_l\right), \tag{11}$$

where $\boldsymbol{X}$ and $\boldsymbol{\omega}$ are discussed below and in Appendix A. Appendix A also includes derivation of equation (11).

[21] Equation (11) produces a linear propagation (accomplished using the derivatives $\partial\Psi/\partial\boldsymbol{\theta}$) of the parameter uncertainty expressed by $(\boldsymbol{X}^T\boldsymbol{\omega}\boldsymbol{X})^{-1}$ to obtain the variance of Ψ. Equation (11) produces values identical to the Sobol' variance in equation (9) under ideal conditions, including linearity of Ψ with respect to the parameters in the range of parameters defined, and use of consistent information on parameters in the two methods. Linearity is generally more nearly approached as the range of parameter values becomes smaller.

[22] For comparison with Sobol', $\boldsymbol{X}$ and $\boldsymbol{\omega}$ of equation (11) need to include only terms associated with prior information on each parameter (see explanation in Appendix A). In this circumstance, $\boldsymbol{X}$ is a $k \times k$ identity matrix. In addition, for the prior information to be consistent with the parameter ranges defined for Sobol' and the even distribution of sampling points produced by a uniform or Sobol' sampling, each diagonal of ω needs to equal the reciprocal of the variance of a uniform distribution defined using $\theta_{j,min}$ and $\theta_{j,max}$ [e.g., *Mood et al*., 1974] (see also Appendix B). Under these conditions, the equation for the variance for a three-parameter problem would equal:

$$V_L(\Psi)=\left|\frac{\partial \Psi}{\partial \theta_1}\right|_l^2 s_1^2+\left|\frac{\partial \Psi}{\partial \theta_2}\right|_l^2 s_2^2+\left|\frac{\partial \Psi}{\partial \theta_3}\right|_l^2 s_3^2, \tag{12}$$

with s_1^2, s_2^2, and s_3^2 a priori parameter variances. The first-order sensitivity measure for the jth parameter is then calculated at each sample point as [e.g., *Helton*, 1993; *Borgonovo*, 2007; *Saltelli et al*., 2008, and others]:

$$S_{L1}^j=\frac{\left|\frac{\partial \Psi_l}{\partial \theta_j}\right|_l^2 s_j^2}{V_L(\Psi)}, \tag{13}$$

[23] Defining the parameter variance s_j^2 to be consistent with a uniform distribution with limits $\theta_{j,min}$ and $\theta_{j,max}$ makes equation (13) equivalent to an equation presented by *Sobol' and Kucherenko* [2010, proof of theorem 1] if their parameter normalization to an interval of 0–1 is considered. Equation (11) does not include terms such as presented by *Sobol' and Kucherenko* [2010] that could be used to approximate parameter interactions relative to the performance criteria.

[24] The cost for calculating S_{L1}^j at N_l sample points (parameter sets) is $N_l\,(k+1)$. $N_l = 1$ provides results at one location, and commonly local statistics are considered only to be calculated for $N_l = 1$. For DELSA, we consider opportunities provided by $N_l > 1$, where we either distribute the samples uniformly or using the quasi-random Sobol' sequence. For either sampling method, in DELSA the local sensitivity indices are calculated throughout parameter space. Generally, the full frequency distribution of the sensitivity measures is reported. However, for some comparisons to Sobol' measures the median value is reported. In addition, sets of individual local values for each parameter are evaluated for the field case.

2.3. Connection Between DELSA, Method of Morris, and RSA

[25] In the Method of Morris (MoM) [*Morris*, 1991] and the closely related Elementary Effects (EE_j) [*Saltelli et al*., 2008] and Latin Hypercube variant One-At-Time (LH-OAT) method of *van Griensven et al*. [2006], model runs are conducted for a base set of parameter values and for a sequence of sets for which each has one parameter value changed at a time. The fundamental difference is that MoM, EE_j and LH-OAT focus on providing global sensitivity measures.

[26] MoM produces two global statistics: one measures the overall parameter importance and the other provides an overall global measure of importance variability [see *Saltelli et al*., 2008, p. 117]. One would expect the first MoM statistic to identify the same important and unimportant parameters as the mean or median of the DELSA first-order statistics. However, unlike the DELSA first-order statistic, the MoM statistic is not expected to have values that are numerically identical to Sobol' values under any condition. The MoM parameter importance variability measure would tend to be large when the standard deviation of the DELSA distribution is large, but MoM does not reveal the distribution of local importance across parameter space, for which DELSA was developed. This is consistent with the general use of large parameter value changes in MoM instead of the small values used to evaluate the derivatives of equation (10). Also, in MoM and EE_j the sequence of one-at-a-time parameter values is sometimes cumulative instead of starting each parameter value change at the same point in parameter space. That is, in MoM first one parameter changes value. From that new location, another changes value, and so on. One parameter may be changed many times before proceeding to the next parameter. In contrast, in DELSA the goal of obtaining local derivatives means that each one-at-a-time parameter value change occurs from the same point in the parameter space, the changes are intentionally small, and typically one change occurs. In this regard, DELSA is similar to LH-OAT, but the latter was used to produce global measures.

[27] Distributions of sensitivities like those produced by DELSA have to our knowledge not been produced in applications of the other methods. [*Saltelli et al*., 2008, p. 128] suggests that MoM "overcomes the limitations of a local

derivative-based approach in that it attempts to explore the whole parameter space." DELSA seeks to accomplish this same goal using a different approach that provides more detail at the local scale. Experience will be needed to better understand the relative insights and utility of these two methods.

[28] DELSA also has some similarities to regional sensitivity analysis (RSA) [*Hornberger and Spear*, 1981; *Freer et al*., 1996]: RSA evaluates the frequency distribution of a model prediction metric across the parameter space, whereas DELSA evaluates the frequency distribution of the gradient of a model prediction metric with respect to the parameter values.

3. Experiments for a Synthetic Test Case

3.1. Simple Nonlinear Reservoir Model

[29] The Sobol' and DELSA sensitivity analysis methods are first applied to the following simple nonlinear reservoir model with two model parameters

$$\frac{dS}{dt} = p(t) - q(t), \tag{14}$$

where S (mm) is model storage, t (day) is time, $p(t)$ (mm d^{-1}) is precipitation (model input), which is kept zero in this case, and $q(t)$ (mm d^{-1}) is drainage (model output).

[30] Drainage is calculated as

$$q(t) = K\left(\frac{S(t)}{S_{sc}}\right)^{c}. \tag{15}$$

[31] This can be seen as the flux resulting from gravitational drainage when the Clapp-Hornberger model of unsaturated hydraulic conductivity [*Clapp and Hornberger*, 1978] is used in Richards' equation. K (mm d^{-1}) and c (−) are two model parameters for which sensitivity indices are analyzed. The scaling parameter S_{sc} (mm) is a constant and is assigned an arbitrary value of 1000 mm. The initial model storage $S(t = 1)$ is set to 600 mm and the model simulation is composed of six hundred 1 h time steps (25 days long). The parameter range for K is defined as [1,1000] and for c as [1,4]. Finally, the performance index Ψ is defined as the mean storage S over time for most results, though results for 10 hourly averaged storages are also presented. The parameter ranges and initial conditions are defined such that the mean storage has a wide spectrum of possible values.

[32] Equation (14) is solved using a fixed-time step, implicit Euler numerical scheme by implementing the Newton-Rapshon method. This numerical implementation is unconditionally stable and avoids numerical artifacts [*Clark and Kavetski*, 2010].

[33] The time series of simulated model storage S for different combinations of parameters K and c with the scaling parameter S_{sc} kept constant are shown in Figure 1. The

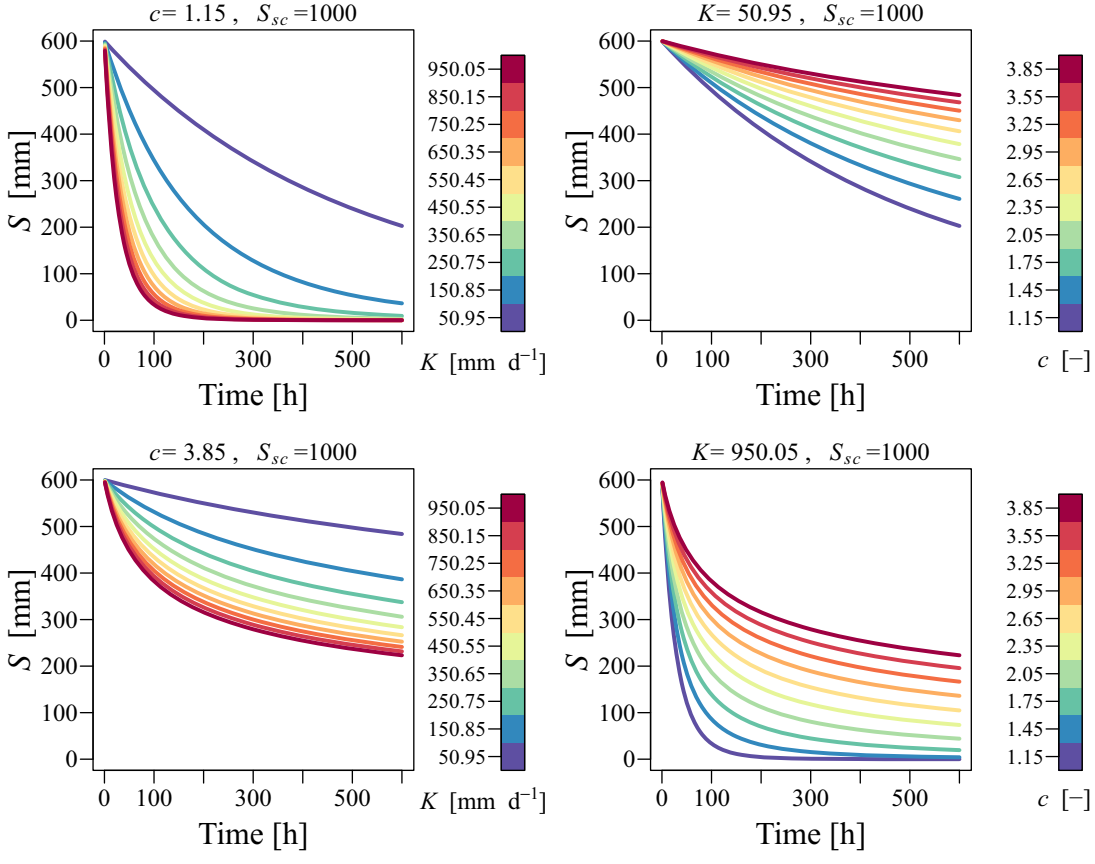

Figure 1. Results of dry down experiment with the simple nonlinear model. Colored lines indicate various values of parameters K and c.

dynamics of this simple model form part of the more complex conceptual hydrological models (further presented in section 4): drydown from "field capacity," which is a typical saturation degree for the Lasnenville catchment (section 4.2). Figure 1 shows that higher values of K yield proportionally higher outflow and model storage decreases faster. Additionally, when the parameter c also has lower values, the reservoir storage empties completely in 600 h. In contrast, for lower values of K and higher values of c, model storage slowly decreases.

[34] The time-averaged model storage, which is used as a model performance index Ψ (equation (1)), is shown in Figure 2 on a regular grid for 100 parameter combinations of K and c, in which the storage behavior corresponds with the simulations in Figure 1. For the defined parameter ranges of K and c, Ψ values range between 32 and 534 mm.

3.2. Comparison of Sobol' and DELSA Total Variances

[35] To compare the Sobol' and DELSA standard deviations (the square root of the total variances), values are calculated for the 100 points of the 2-D parameter space, for which Ψ values are reported in Figure 2. For DELSA, this is accomplished by calculating the local $\sqrt{V_L(\Psi)}$ of equation (11) at the midpoints at each location. For Sobol', $\sqrt{V(\Psi)}$ (equation (2)) is calculated for the 100 parameter subdomains that surround the 100 locations. The resulting DELSA and Sobol' standard deviations are shown in Figures 3 (top) and 4; results produced by the two methods are very similar. Sobol' values are slightly larger, as expected given that the local methods used here do not account for parameter interactions related to Ψ. As mentioned in the methods section, *Sobol' and Kucherenko* [2010] suggest terms for including these interactions, but they have not been included in this work. For this situation, the small difference between Figures 3 (top) and 4 suggests this omission has little consequence.

[36] Using Sobol' to calculate subdomain values is very computationally expensive: 2,000,000 model runs are needed to produce Figure 4, since the computation costs for the Sobol' variance in equation (9) require $N_S k$ model runs.

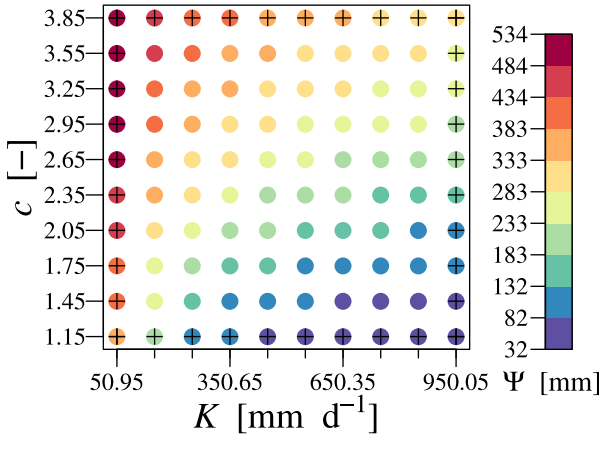

Figure 2. Sensitivity of performance index Ψ (mean storage S) of the simple nonlinear model to parameters K and c. Plusses indicate parameter combinations for which time series are shown in Figure 1.

Such a large number of model runs is usually possible only for simplified problems such as presented here and the subdomain Sobol' runs are only attempted to demonstrate the similarity between local and global variance estimates under close to ideal conditions.

[37] Comparison of the local standard deviations of Figure 3 (top) to the local gradients (equation (10)) used by DELSA (Figure 3, bottom) shows how the patterns of the derivatives mirror the patterns of the standard deviations. The magnitudes of the local gradients differ, because of the defined parameter ranges. Derivatives are generally smaller for K, because the K values are relatively large (derivatives range from 0.0 to 3.0; K values range from 50.95 to 950.95). Derivatives are generally larger for c, because the c values are relatively small (derivatives range from about 25 to 175; c ranges from 1.15 to 3.85). Therefore, the local values need to be scaled to obtain comparable measures. In equation (11), the derivatives in $d\Psi/d\boldsymbol{\theta}$ are scaled by the prior parameter variances in $\boldsymbol{\omega}$ to obtain standard deviations plotted in Figure 3 (top).

[38] For DELSA, the matrix $\boldsymbol{\omega}$ of equation (11) was defined to be consistent with the parameter value ranges used by Sobol'. DELSA values comparable to Sobol' values calculated for 10 × 10 parameter subdomains require that the prior parameter variances used in DELSA be determined using a uniform distribution, yielding $s_K^2 = 28.839$ and $s_c^2 = 0.087$ (see Appendix B).

3.3. Comparison of Sobol' and DELSA First-Order Sensitivity Indices

[39] Sobol' and DELSA first-order sensitivity indices are calculated for the 100 subdomains for which the performance index and standard deviations are shown in Figures 2–4. For Sobol', equation (3) is applied to each parameter subdomain; for DELSA, equation (13) is applied at the middle point of each subdomain. The results are shown in Figure 5, and suggest that the methods provide similar sensitivity indices across the defined parameter space. Similar to the Sobol' calculation of the total variance, these simulations suggest that calculation of the Sobol' sensitivity measure distributed in parameter space is computationally expensive. Here, the calculation costs for the Sobol' first-order index are N_S $(k+1)$ times 100 subdomains (=3,000,000), while for the local equivalent this is N_l $(k+1)$ (=300). This suggests that if the distributions produced by DELSA are of interest, obtaining them using Sobol' is likely to be impractical for most problems.

[40] The traditional approach in the Sobol' method is to calculate one sensitivity measure for each parameter for the whole parameter space. The Sobol' S_1 indices for K and c are shown in Figure 6 (top) and they have approximately the same sensitivity of 0.50 and 0.48, respectively. Figure 6 (middle) presents the results of Figure 5 (top) showing the cumulative frequency distributions of the Sobol' indices for the 100 parameter subdomains. Of interest is that the central values differ from the Sobol' global values: the median values are $S_1 = 0.33$ for K and $S_1 = 0.81$ for c.

[41] To understand differences between the traditional way of calculating the Sobol' method (Figure 6, top) and the local application of the Sobol' method (Figure 6, middle), we decompose the total variance $\hat{V}(\Psi)$ into its constituents for each of the 100 parameter subdomains. These are

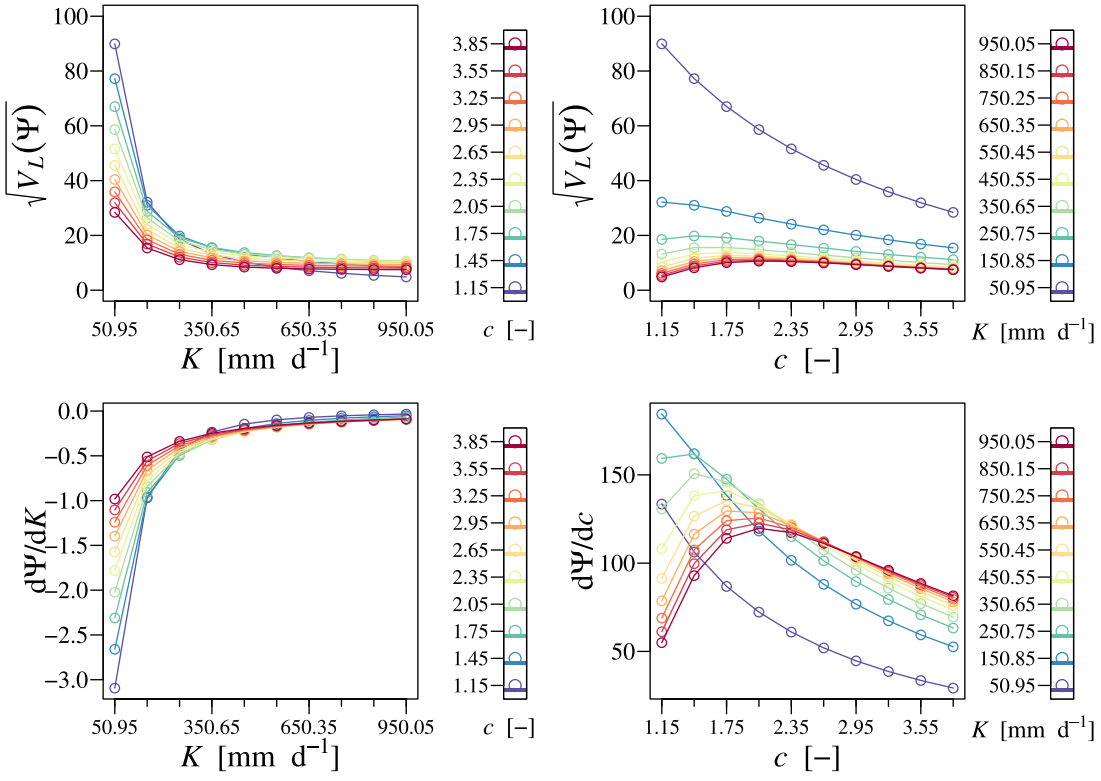

Figure 3. (top) Prior model prediction standard deviation ($\sqrt{V_L(\Psi)}$) calculated for values of parameters K and c at 100 points across the parameter space as indicated in Figure 2. (bottom) Local gradients of the parameters K and c to model prediction Ψ calculated for values of parameters K and c at the same 100 points. These two graphs plot the same 100 values in two ways to show how the standard deviations ($\sqrt{V_L(\Psi)}$) are related to the sensitivities.

labeled $(\Psi_i - \overline{\Psi})^2$ in Figure 7, where i indicates individual parameter locations and $\overline{\Psi}$ is the average over the entire parameter space. Note that calculating the mean of all $(\Psi_i - \overline{\Psi})^2$ terms approximately equals the total variance of Ψ, which is 17,091 mm. It is clear that different parts of the parameter domain produce different contributions to $\hat{V}(\Psi)$. For example, regions with higher c and lower K values produce large contributions to $\hat{V}(\Psi)$ and lower c and higher K values also have significant contributions. Over half of the parameter domain contributes little to the variance. These results indicate that while the first-order Sobol' sensitivity indices are calculated across the whole parameter domain, only certain subregions significantly contribute to the sensitivity indices. This explains why the results between the traditional (global) and the distributed Sobol' indices differ.

[42] The sample size needed for reliable results is important because it governs the computational demands of any

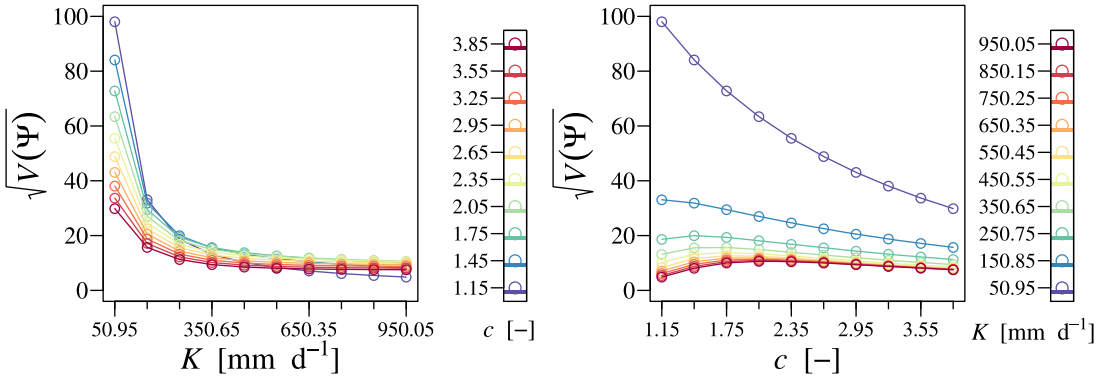

Figure 4. Standard deviation of the model prediction Ψ across the parameter space calculated using the Sobol' variance for 100 parameter subdomains with 10,000 sample size each. Delineated parameter subdomains are shown in top plot of Figure 5.

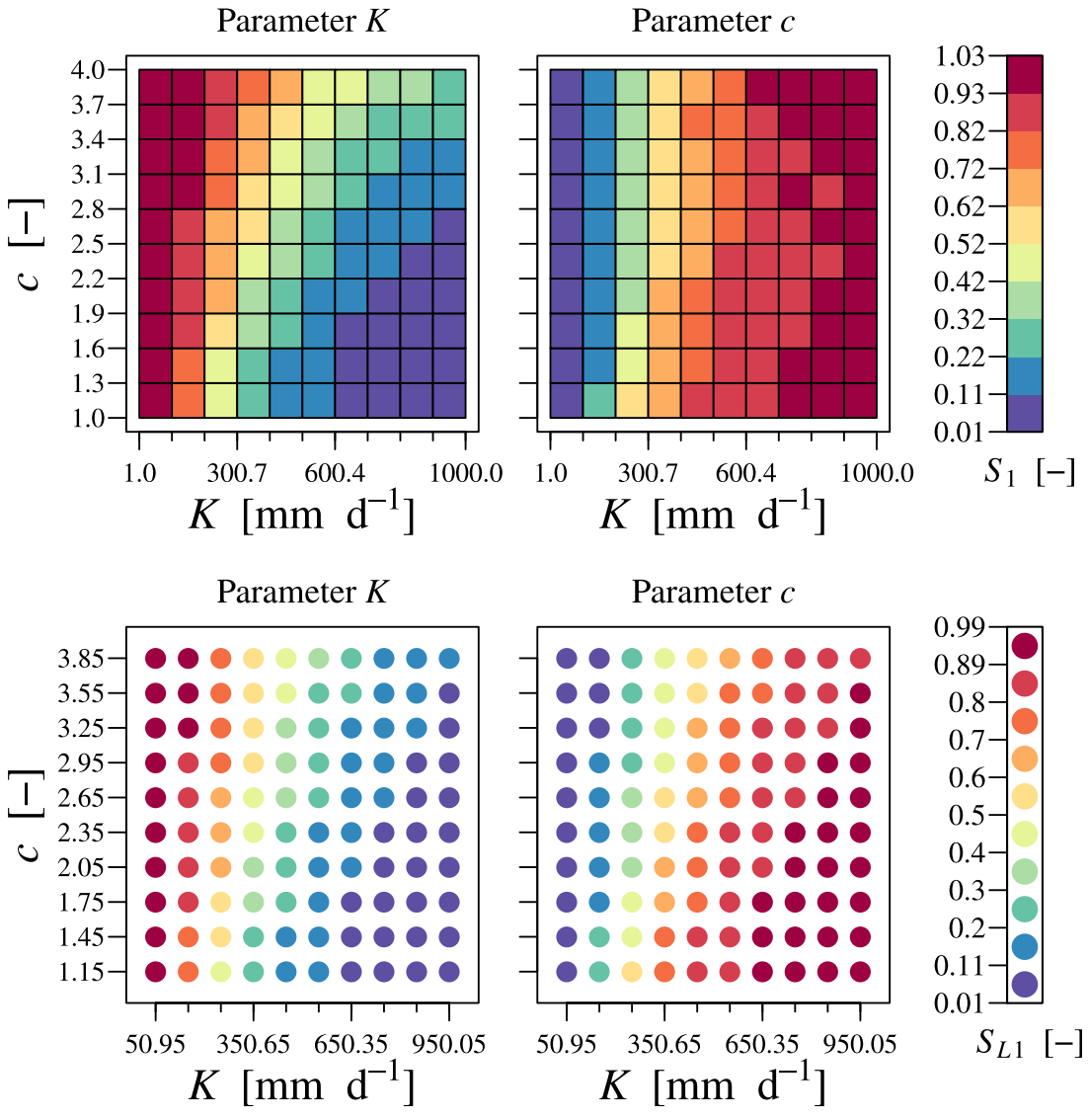

Figure 5. Comparison of first-order sensitivity statistics of (top) the global method with Sobol' sampling (S_1) and (bottom) the DELSA method (S_{L1}) for 100 parameter subdomains with 10,000 sample size each.

method. The DELSA cumulative frequency distributions of the local first-order sensitivity index (S_{L1}) for different sample sizes are shown in Figure 6 (bottom). Sampling on a regular grid as in Figure 5 does not generalize easily when considering different sample sizes. Thus, the quasi-random Sobol' sequence is used to define the samples used for Figure 6 (bottom). The DELSA approach provides a full distribution of the first-order indices and we can easily obtain a number of different statistical properties. For example, the results in Figure 6 (bottom) suggest that parameter c is more influential in much of the parameter space (with median of S_{L1} equal to 0.75) than the parameter K (with median of S_{L1} equal to 0.25). DELSA attains stable results with only 300 model runs ($N_l = 100$), with no significant advantage obtained from using a higher N_l.

[43] Finally, the effect of sampling uncertainty on the first-order sensitivity indices S_{L1} and S_1 is presented in Figure 8, where the sampling variability is obtained from 1000 bootstrap samples. The median values for $N_S = 10000$ in Figure 8 are the indices presented in Figure 6 (top). When compared with the DELSA method, the Sobol' method required much larger N_S (>1000) to provide meaningful results within the expected interval [0,1]. The sensitivity values of Sobol' and DELSA are not expected to converge to the same value because of the averaging effects in the global methods, as discussed using the total variance decomposition (Figure 7).

3.4. Time Varying DELSA

[44] For transient problems, DELSA can be plotted over time for sampled parameter sets to understand changes in dominant model controls as suggested, e.g., by *Cloke et al.* [2007] and *Herman et al.* [2013a]. For the synthetic test case, consider mean storage calculated over 60 consecutive

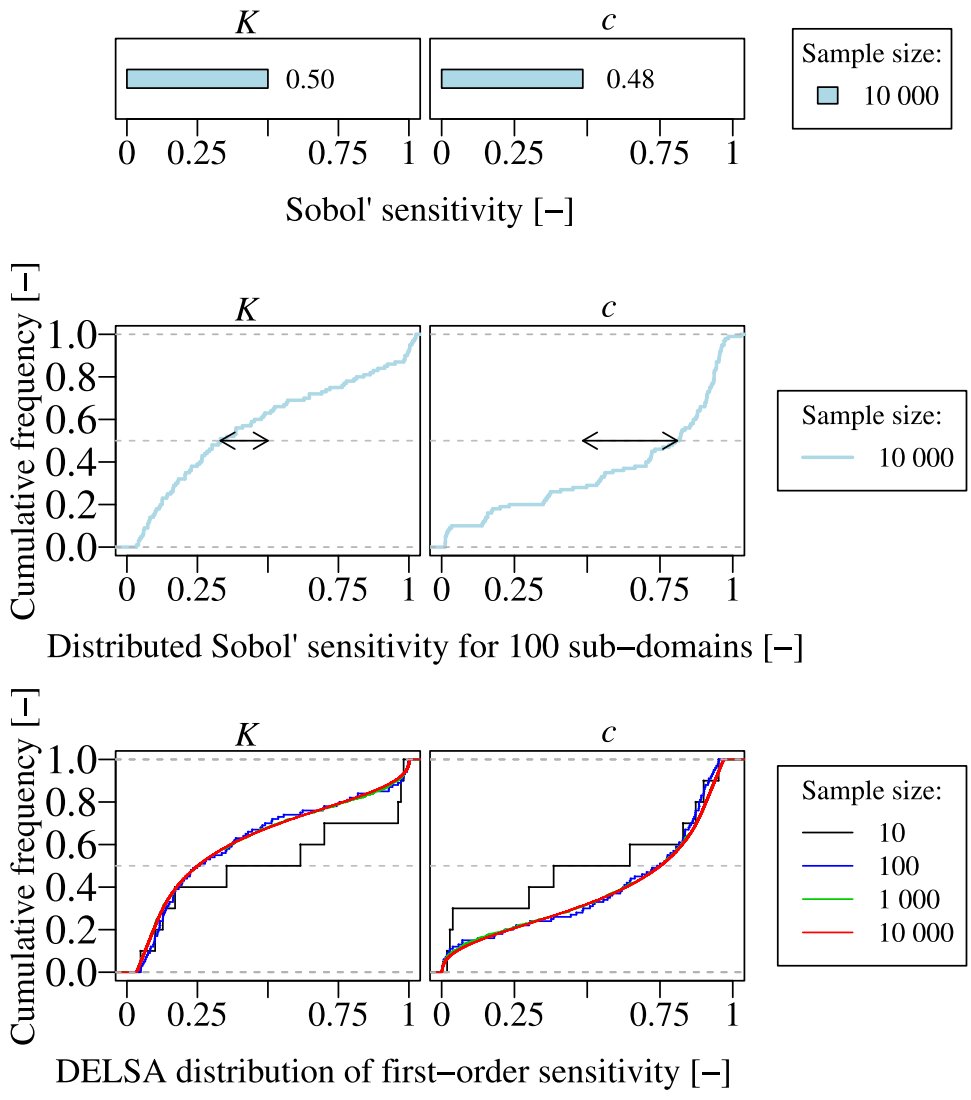

Figure 6. Sensitivity indices for the simple nonlinear model with two parameters K and c. (top) Bar chart of the first-order Sobol' global sensitivity indices (S_1) for K and c across the whole parameter space. (middle) Cumulative frequency distributions of the distributed Sobol' indices for 100 subdomains. The arrows show the difference from the global Sobol' index presented in the top plot, which are thought to result from model nonlinearity. The nonlinearity is apparent in the distribution of the DELSA statistic; for linear models the same value would be expected throughout parameter space. (bottom) Cumulative frequency distributions of the DELSA method (S_{L1}) obtained for different sample sizes (N_l) derived using the quasi-random Sobol' sequence, showing that only 100 samples provide accurate results for this problem.

10 h periods. We illustrate this in Figure 9 for the simple model using five combinations of parameters K and c: the four corner points and center point of Figure 2.

[45] DELSA values are time invariant for the lower K values and temporally variable for higher K values. For some parts of the parameter space (e.g., $K = 450.55$ and $c = 2.65$), the ranking of the parameter importance changes over time from one parameter to the other. In a given problem, observed transient system performance could likely be used to identify which DELSA results are most relevant, in a similar manner as dynamic identifiability analysis (DYNIA) proposed by *Wagener et al.* [2003].

4. Experiments for Real-World Case Studies

4.1. Alternative Hydrologic Model Structures

[46] To understand the performance of DELSA in realistic problems, we consider a set of conceptual hydrological models within the Framework for Understanding Structural Errors (FUSE) [*Clark et al.*, 2008]. Multiple models are considered because of their utility in practice [e.g., *Ye et al.*, 2010; *Clark et al.*, 2011; *Foglia et al.*, 2013]. Alternative models allow evaluation of alternative hypotheses about system dynamics, and these alternatives can affect which parameters are important and unimportant. For one model a parameter may be insensitive over most of the

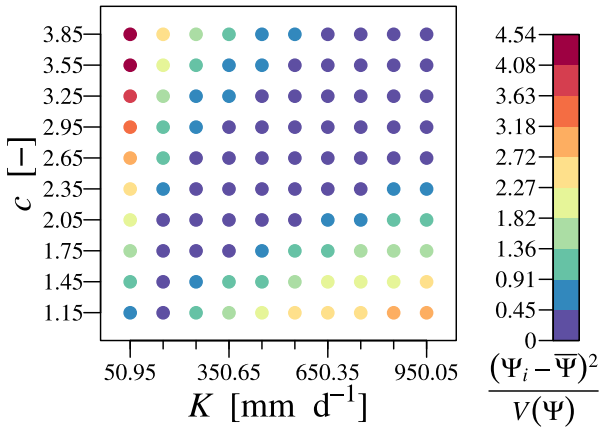

Figure 7. Decomposition of the total variance of the model performance index Ψ into its constituents at the 100 parameter space locations. Each constituent is divided by the global variance calculated using Sobol' for combinations of parameters K and c.

parameter space, while in another the same parameter can be sensitive over much of the parameter space. The less computationally demanding DELSA makes evaluation of many models more convenient.

[47] In this study, we implement five FUSE models (FUSE-016, FUSE-014, FUSE-160, FUSE-072, and FUSE-170) analyzed by *Clark et al*. [2011]. These models represent different modeling decisions, which include: (1) the choice of state variable in the unsaturated and saturated zones and (2) the choice of flux equations describing the surface runoff, vertical drainage between soil layers, base flow, and evapotranspiration. For a detailed description of the hydrological models and their model equations, we refer to *Clark et al*. [2008, 2011]. Here, we provide a brief description.

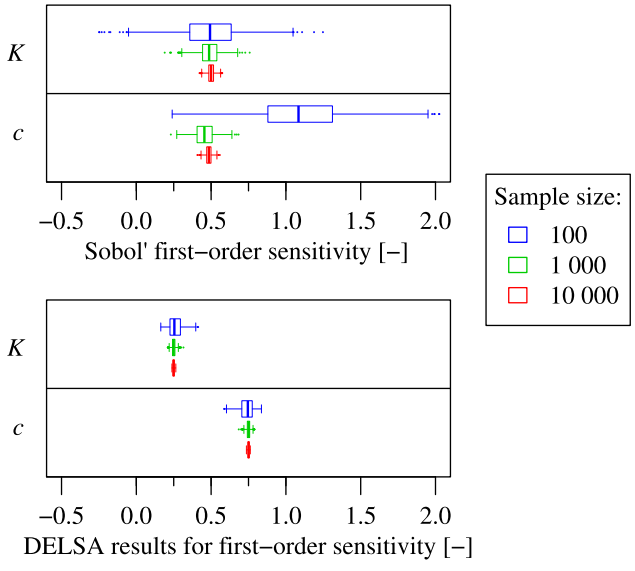

Figure 8. Sampling uncertainty for the nonlinear model with two parameters K and c. (top) Effect of the sample size (N_S) on the Sobol' sensitivity index (S_1). (bottom) Effect of the sample size (N_l) on the median of the DELSA method (S_{L1}). The uncertainty estimates are obtained by bootstrapping (resampled 1000 times). The vertical bold line in the boxplot is the median, the body of a boxplot shows the interquantile range (Q75–Q25) and the whiskers represent the sample minima and sample maxima, unless the extreme value occurs further than 1.5 times the interquartile range from the box, in which case the outlier is shown by a dot.

[48] The models FUSE-016, FUSE-014, and FUSE-160 evaluate the choice of state variables used to represent the unsaturated zone and different parameterizations of evapotranspiration. FUSE-016 has a single reservoir; FUSE-014 and FUSE-160 use two cascading reservoirs. With respect to evapotranspiration, FUSE-016 implements a "single-layer" parameterization, FUSE-014 implements a "sequential" parameterization, and FUSE-160 implements a "root weighting" parameterization. All other model components are kept constant, including a single nonlinear groundwater reservoir of unlimited size, the "ARNO/VIC (Variable Infiltration Capacity)" parameterization of surface runoff, and time delay routing using a gamma distribution.

[49] The FUSE-072 model is used to evaluate different vertical drainage parameterizations. FUSE-016, FUSE-014, and FUSE-160 do not allow any vertical drainage when saturation is below field capacity. The vertical drainage in the FUSE-072 model is a nonlinear function of total storage in the unsaturated zone and for the rest FUSE-072 is identical to FUSE-016. The FUSE-170 model is used to evaluate different base flow parameterizations. It uses two parallel linear reservoirs to represent base flow.

[50] The parameters used in the five models are described in Table 1. The effective ranges applied in this study were slightly adjusted from *Clark et al*. [2011] based on our prior knowledge about the Lasnenville catchment (section 4.2). The FUSE parameters PERCRTE, PERCEXP, and MAXWATR_1 can be seen as the equivalent of the simple nonlinear reservoir model (section 3.1) with parameters K, c, and S_{sc}, respectively (equation (15)). Like the simple nonlinear model, the FUSE simulations have an hourly time step, and fixed-step implicit Euler numerical approximation.

[51] The model performance index Ψ (equation (1)) for this problem is defined as the root-mean-square-error (RMSE) between the observed streamflow (q_{obs}) and FUSE streamflow simulations (q_{sim}):

$$\Psi = \mathrm{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(q_{obs,t} - q_{sim,t})^2}. \quad (16)$$

[52] This measure tends to emphasize high flows. Sensitivity analysis results can change when using different performance indices.

4.2. Data

[53] The sensitivity analyses using the FUSE models are carried out for a 10 year period from 1 October 1998 to 30 September 2008 for the Lasnenville catchment located in the Belgian Ardennes (Western Europe). The Lasnenville catchment is a medium-sized catchment (200 km^2) located upstream of the Lasnenville streamflow gauge, which is operated by the Hydrological Service of the Walloon

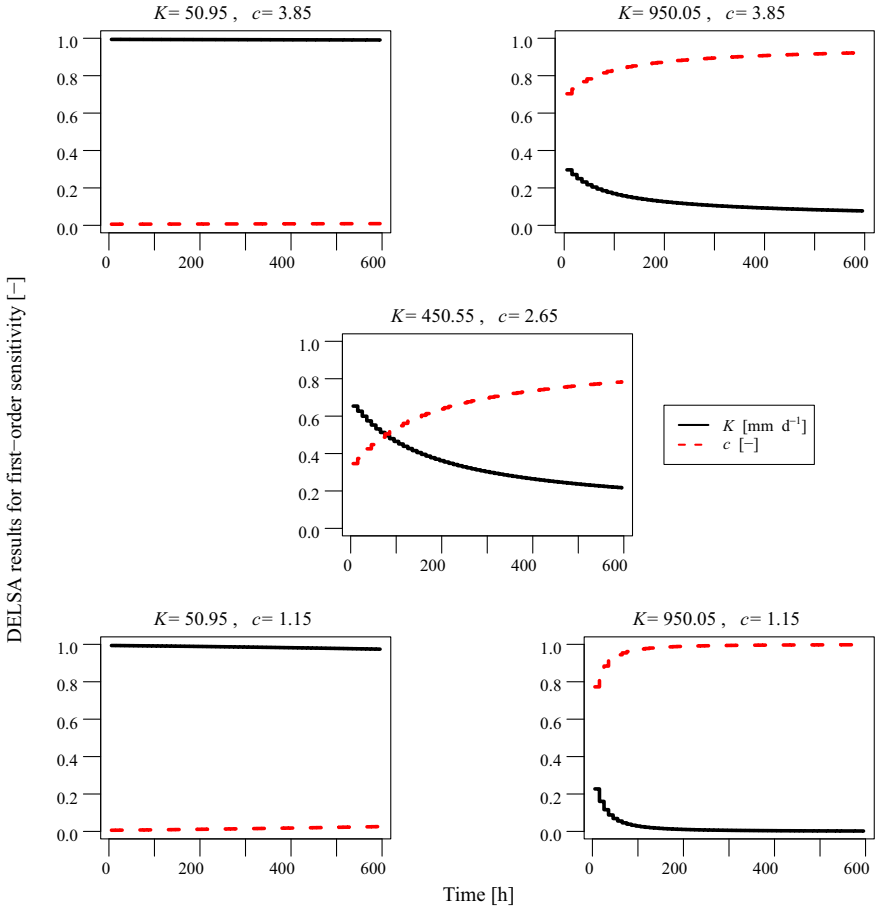

Figure 9. Changes in DELSA sensitivity over time for the drydown experiment. DELSA sensitivity (S_{L1}) is calculated for 60 consecutive 10 h periods at five combinations of the K[mm d^{-1}]-c parameter space.

Region of Belgium (MET-SETHY). The meteorological data are obtained from the Vielsalm FLUXNET site [*Aubinet et al*., 2001; *Papale et al*., 2006; *Reichstein et al*., 2005], which is located at the border of the Lasnenville catchment. The cumulative annual precipitation is about 1000 mm and mean annual temperature 7.5°C. The land use is approximately half mixed-forest and half agriculture. The climate conditions can be classified as rain fed with ephemeral snow in winter and the runoff regime is highly variable with low summer discharges and high winter discharges [e.g., *Driessen et al*., 2010; *Hazenberg et al*., 2011; *Rakovec et al*., 2012].

4.3. Comparison of Sobol' and DELSA

[54] The sensitivity indices of all model parameters are shown in Figure 10 for sample sizes $N_S = 1000$ and $N_I = 1000$, respectively. The same sample size was used for both Sobol' and DELSA so that use of different sample sizes would not confuse intercomparison of results. Because the FUSE models have between 11 and 14 parameters and the calculation costs are $N_S\,(k + 1)$ and $N_I\,(k + 1)$, respectively, the number of model runs is between 12,000 and 15,000, depending on a model structure. For DELSA, the distribution of median values is shown.

[55] Results suggest that, for all models for DELSA and Sobol', parameters TIMEDELAY and AXV_BEXP are the most important (see Table 1 for a description of each parameter). This makes sense, because these two parameters significantly affect the magnitude and timing of the simulated streamflow in the RMSE. Other parameters are moderately important to one or more models. For example, FRACTEN is the third most sensitive parameter for FUSE-016 according to both methods, which is consistent with how evaporation and unsaturated zone are simulated [*Clark et al*., 2011]. For the FUSE-072 model, PERCEXP is somewhat important because vertical drainage from the unsaturated zone to the saturated zone is simulated even when the upper zone storage falls below field capacity. FUSE-160 uses two parallel linear reservoirs to represent base flow, in contrast to the single nonlinear reservoir used in all other FUSE models, and there are three other sensitive parameters: MAXWATR_1, PERCFRAC, and QBRATE_2B (see Table 1 for parameter description).

[56] For all five models, the DELSA approach provides sensitivity indices, which correspond well to the global Sobol' method.

[57] The DELSA method yields much smaller sampling uncertainty than the global Sobol' method for the same number of model runs. This is shown in Figure 10 by the error bars accounting for the 95% confidence interval: for a sample size of 1000, the DELSA error bars are much

Table 1. Parameters of the FUSE Models Used in This Work

Parameter Name	Description	Units	Lower Limit	Upper Limit
MAXWATR_1	Maximum storage in the unsaturated zone	mm	50	500
MAXWATR_2	Maximum storage in the saturated zone	mm	25	250
FRACTEN	Fraction total storage as tension storage	–	0.05	0.95
FRCHZNE	Fraction of tension storage in the primary zone (unsaturated zone)	–	0.05	0.95
FPRIMQB	Fraction of free storage in the primary reservoir (saturated zone)	–	0.05	0.95
RTFRAC1	Fraction of roots in the upper soil layer	–	0.05	0.95
PERCRTE	Vertical drainage rate	mm d^{-1}	0.01	1000
PERCEXP	Vertical drainage exponent	–	1	20
PERCFRAC	Fraction of drainage to tension storage in the lower layer	–	0.05	0.95
FRACLOWZ	Fraction of soil excess to lower zone	–	0.05	0.95
BASERTE	Base flow depletion rate for the single reservoir	mm $d^{—1}$	0.001	1000
QB_POWR	Base flow exponent	–	1	10
QBRATE_2A	Base flow depletion rate for the primary reservoir	day^{-1}	0.001	0.25
QBRATE_2B	Base flow depletion rate for the secondary reservoir	day^{-1}	0.001	0.25
AXV_BEXP	ARNO/VIC "*b*" exponent for the surface runoff	–	0.001	3
LOGLAMB	Mean of the log-transformed topographic index distribution	m	5	10
TISHAPE	Shape parameter defining the topographic index distribution	–	2	5
TIMEDELAY	Routing parameter equal to the time delay in runoff	day	0.01	2

narrower. This was also analyzed for the two most sensitive parameters TIMEDELAY and AVX_BEXP using sample sizes of 100, 1000, and 10,000 (not shown). While 100 model runs provided consistent sensitivity indices using the DELSA method, Sobol' required many more model runs (preferably $N_S > 1000$) to provide sensitivity indices with equivalently low sampling uncertainty.

[58] In Figure 10, the Sobol' values and median DELSA values are sometimes different, though the same important and unimportant parameters are identified. Such differences are expected and are discussed in section 3 of this work.

[59] Figure 11 compares the full distribution of parameter sensitivity provided by the DELSA method to the global Sobol' model. Results are shown for seven parameters common to all five FUSE models. While Figure 10 shows only the median sensitivity values for DELSA, Figure 11 shows the full empirical frequency distribution. Sensitivity indices at the cumulative frequency of 0.5 correspond to the values shown in Figure 10. The DELSA results allow for more detailed evaluation. For example, consider two parameters: FRACTEN in FUSE-016 and PERCEXP in FUSE-072. Both have a first sensitivity index of 0.1 estimated by the global Sobol' method, but inspection of the DELSA results shows that parameter PERCEXP in FUSE-072 has a sensitivity index >0.75 for about 20% of the sampling locations. This means that for about 20% of the parameter value samples, PERCEXP is nearly as important as the two most important parameters, TIMEDELAY and AVX_BEXP. On the other hand, the FRACTEN parameter in FUSE-016 hardly ever exceeds a DELSA sensitivity of 0.5, even though the median shows a sensitivity of 0.1. Inspection of the DELSA results also suggests that each parameter is insensitive for about 20% of the parameter sets. Inspection of some model runs (not shown) for which the sensitivity of the TIMEDELAY and AVX_EXP parameters equals zero showed that the FRACTEN parameter became very important. For these models, the simulated response was either much more or much less flashy than the data suggested. Further evaluation is beyond the scope of the present work, but even these initial results suggest the insight available through multiscale evaluations of sensitivity such as that provided by DELSA.

[60] One interesting question is how global sensitivity measures are affected by results in regions of parameter space for which the model provides a poor representation of actual conditions. *Pappenberger et al*. [2008] and others have approached this issue by only performing global sensitivity analysis on the model simulations with "acceptable" model performance—simulations that *Pappenberger et al*. [2008] considered to be "behavioral." Figure 12 illustrates values of the DELSA statistic for each parameter value for model FUSE-016, plotted against the RMSE performance metric. Here, we make two observations. First, there is a small number of outlier parameter sets, denoted as triangles in Figure 12, for which there is high sensitivity for FRACTEN, AXV_BEXP, and lower sensitivity for parameter TIMEDELAY. These parameter sets may provide unrealistic simulations, as indicated by the high RMSE statistic. These parameters sets all have values of TIMEDELAY that are less than 0.05; the range of TIMEDELAY was defined as 0.01–2, and the small values would not be expected for a basin this large (200 km^2). These results could be used to understand why model performance is so anomalous, or to omit such unrealistic results from the analysis. Second, parameter sets that have the highest values of sensitivity in the TIMEDELAY also tend to have poor model performance. To the extent that large values of RMSE indicate an unrealistic model, this may suggest that for models of most interest the dominance of parameter TIMEDELAY may be exaggerated by the global approach. Figure 12 also shows the difficulties faced by using traditional $N_l = 1$ sensitivity analysis for these models. Such analyses are likely to identify MAXWATR_2, PERCRTE, and perhaps PERCEXP as insensitive, but results for other parameters are likely to be erratic. Investigation of whether spurious or realistic results account for this is an interesting question not pursued here.

5. Discussion

[61] The results of the simple (synthetic) and complex (real world) problem suggest potential utility of the

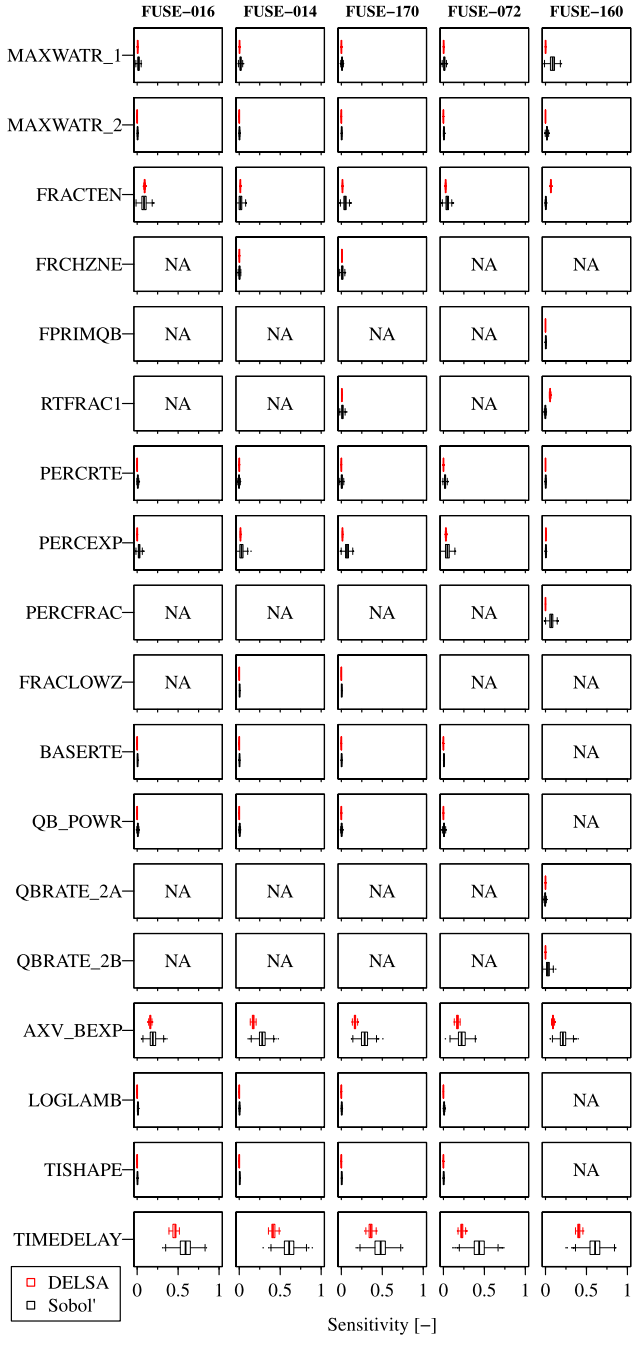

Figure 10. Comparison of the DELSA (S_{L1}) and Sobol' (S_1) sensitivity indices for five FUSE model structures using $N_S = 1000$ and $N_l = 1000$. The sampling uncertainty (bootstrapping with 1000 times resampling) is indicated by boxplots.

DELSA method. Advantages include convergence with relatively small values of N_l, and detailed characterization of variations in sensitivity over the defined range of parameter values.

[62] The local methods considered in this work are expected to perform well when performance metrics are smooth functions of the parameters. Simulated results can be nonsmooth for a number of reasons. These include spurious solutions of the type discussed by *Clark and Kavetski* [2010] and *Kavetski and Clark* [2010], and can also include small oscillations produced by otherwise valid solution schemes (for example, random walk and method of characteristic solutions for transport often produce small oscillations) [*Mehl and Hill*, 2001]. In the presence of irregularities and small oscillations, the small parameter change used to calculate derivatives in local methods can produce sensitivity measures that indicate greater parameter importance than actually occurs (also noted by, for example, *Kleijnen and Helton* [1999] and *Sobol' and Kucherenko* [2010]). This is likely the reason that in *Li et al*. [2013] local methods indicated greater importance for some parameters than did the global methods considered.

[63] In DELSA, such difficulties will not dominate a parameter importance measure as long as the irregularities occupy little of the parameter space. A few large values will clearly be exceptional when the DELSA distributions are plotted, as for example Figure 11 (bottom) or in Figure 12, and can be suitably ignored. Indeed, DELSA provides an opportunity to identify and possibly fix such model irregularities if they are deemed to be unrealistic.

[64] If the irregularities are pervasive and oscillate around a valid solution, DELSA will perform poorly and function smoothing or use of global methods is advised. Often, such irregularities are known to be typical of a modeling method, as is the case for random walk and the method of characteristics (MOC) transport solutions [see *Mehl and Hill*, 2001]. For many reasons besides sensitivity analysis, it is advisable for modelers to be cognizant of solution characteristics such as small oscillations and other irregularities; such knowledge is needed to determine the utility of model results as well as suitable methods for model development and analysis.

[65] The DELSA method provides a number of intriguing opportunities that we do not pursue in this first paper on DELSA. Here, we briefly mention selected opportunities and significant related literature and issues.

[66] Equation (11) is presented in its general form and Appendix A discusses how observations can be included in $\boldsymbol{X}$ and $\boldsymbol{\omega}$ to emphasize that DELSA has close ties to local sensitivity theory [*Cook and Weisberg*, 1982; *Foglia et al*., 2009]. The DELSA method appears to provide rich possibilities for using a full parameter variance-covariance matrix as required to include observations and thus allow sensitivity analysis on a posteriori distributions, as indicated in Figure 13. In global methods, full matrices were considered by *Saltelli et al*. [2004] but Sobol' evaluations were found to depend on the order of the parameters [*Bedford*, 1998]. *Xu and Gertner* [2011], *Zuniga et al*. [2013], and *Kucherenko et al*. [2012] present methods for accounting for parameter correlation in the context of global methods. To our knowledge, numerically equivalent local methods have not been identified and their development is an interesting topic beyond the scope of the present work.

[67] In this work, we focus on the first-order sensitivity measure. Sobol' statistics of higher order, including total-order statistics, could perhaps be included in DELSA by using the local derivatives [e.g., see *Sobol' and Kucherenko*, 2010].

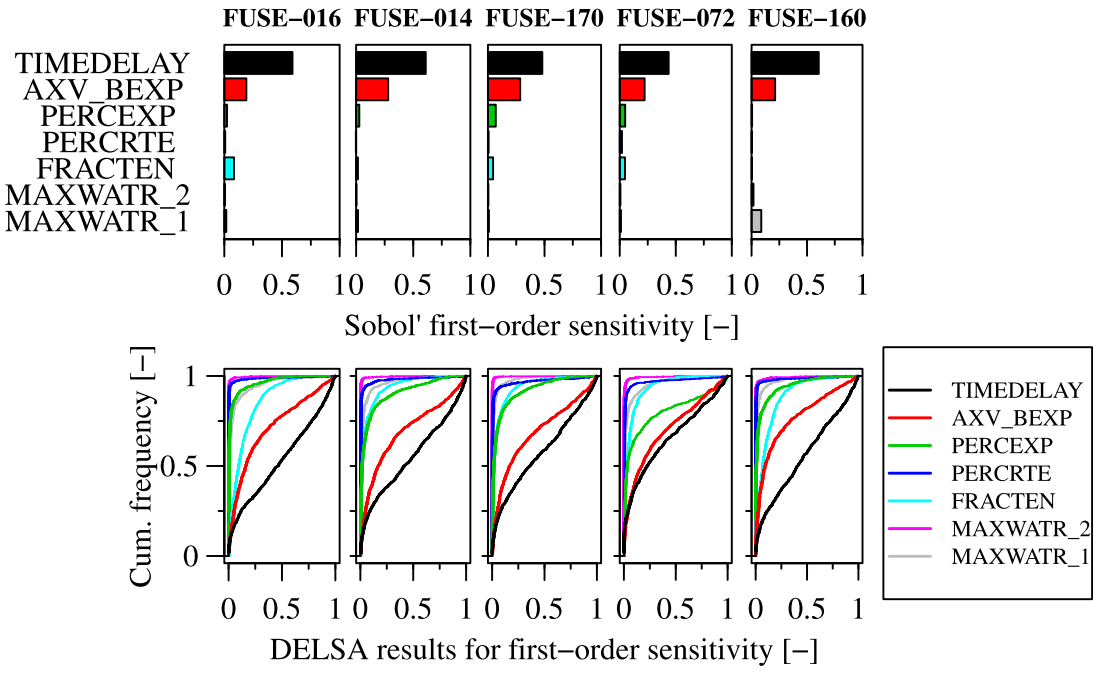

Figure 11. Sensitivity indices for the seven common parameters of the five FUSE model structures using $N_S = 1000$ and $N_l = 1000$. (top) Global Sobol' sensitivity indices (S_1). (bottom) Cumulative frequency distributions of the variability across the parameter space using the DELSA method (S_{L1}).

6. Conclusions

[68] We investigated parameter sensitivity and how it is affected by hydrologic model structure using two very different techniques: a global variance-based method (Sobol') and the proposed Distributed Evaluation of Local Sensitivity Analysis (DELSA) method. In DELSA, parameter sensitivity evaluation is based on gradients of the model performance index with respect to model parameters at multiple points throughout the parameter space and sensitivity metrics are formulated in terms of the distribution of values across the parameter space. Additionally, DELSA is constructed in a way that allows direct comparison to global variance decompositions of Sobol'. The results show that the Sobol' and DELSA methods yield similar results for the problems considered, although the DELSA method provides the full distribution of sensitivity throughout the parameter space at lower calculation costs than is incurred to obtain a single summary measure of sensitivity using Sobol'. For example, in the real-world problem the time delay in runoff is the most important parameter in all models, but DELSA shows that for about 20% of parameter sets it is not important at all and alternative mechanisms and parameters dominate. Moreover, the time delay was identified as important in regions producing poor model fits, whereas other parameters were identified as more important in regions producing better model fits. The detailed information produced by DELSA about model

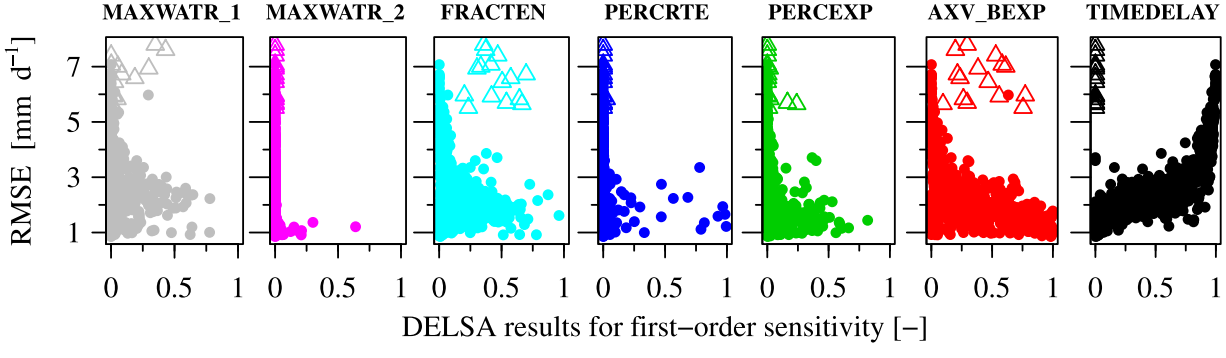

Figure 12. Scatterplot of first-order statistic (S_{L1}) used for DELSA, related to the RMSE performance metric for the FUSE-016 model showing the same parameters as depicted in Figure 11 for $N_l = 1000$. Triangles indicate 15 parameter sets, which are discussed in the text as outliers. Note that simulations with high RMSE are of low hydrological significance. Both AXV_BEXP and TIMEDELAY have high sensitivities, but only in the case of AXV_BEXP do these correspond to low RMSE values.

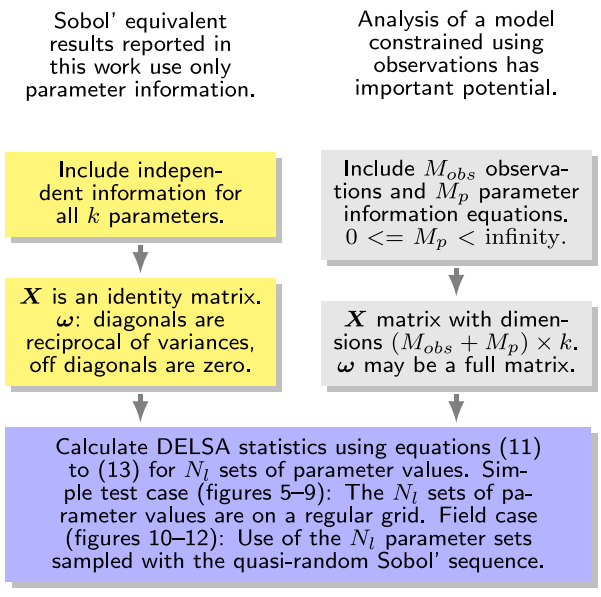

Figure 13. Flowchart showing definition of $\boldsymbol{X}$ and $\boldsymbol{\omega}$ used to calculate DELSA for situations comparable to Sobol' variance analysis (in yellow), as considered in this work. Also shown is how $\boldsymbol{X}$ and $\boldsymbol{\omega}$ are defined in the important situation of having observations with which to constrain the simulation (in gray).

sensitivity provided important insights about how parameter sensitivity varied throughout the parameter space.

[69] This study focuses on introducing the DELSA method and applying the method to reproducible test problems. DELSA is ideally suited to analyze multiple metrics of model behavior [e.g., *Rosolem et al*., 2012] and environmental models with lengthy execution times. The advantage of the multiscale, hybrid local-global DELSA method is that application at multiple points across the parameter space can reveal important parameter subregions, generally undetected by global methods. We anticipate that the DELSA method will become a powerful approach to provide insight about model parameter sensitivity at very low computational cost. We look forward to working with others in the community to apply the DELSA methodology to challenging modeling problems.

Appendix A: Definition of X and ω and Derivation of DELSA Equation (11)

[70] Equation (11) is the prediction total variance, $V_L(\Psi)$, and is composed of two parts.

[71] The first part is the variance-covariance matrix of the parameter vector $\boldsymbol{\theta}$, or $V(\boldsymbol{\theta})=\sigma^2(\boldsymbol{X}^T\boldsymbol{\omega}\boldsymbol{X})^{-1}$, a standard expression stated in many texts, including *Seber and Wild* [1989, p. 191], *Menke* [1989, p. 58], *Draper and Smith* [1998, pp. 129–130], *Tarantola* [2005, p. 36], and *Aster et al*. [2013, p. 224].

[72] $\boldsymbol{X}$ is a matrix with M rows and k columns and $\boldsymbol{\omega}$ is a matrix with M rows and M columns. k is the number of parameters. M depends on the type of problem considered. For model calibration problems with no prior information, M equals the number of observations. If each observation is represented by y_i, the elements of $\boldsymbol{X}$ equal $\partial y_i/\partial\theta_i$. $\boldsymbol{\omega}$ would equal the weight matrix of the observations. If the model calibration problem includes prior information, $M = M_{obs} + M_p$, where M_{obs} equals the number of observations and M_p equals the number of prior information equations. The additional M_p rows of the $\boldsymbol{X}$ matrix often have zeros except for one which identifies the parameter associated with the prior information. However, alternative forms are used to define, for example, first-order Tikhonov regularization. For comparison with Sobol', no observations are used and $M = N_p$. $\boldsymbol{X}$ is a $k \times k$ identify matrix, and $\boldsymbol{\omega}$ is diagonal and contains the reciprocal of the variances of the parameters (see Appendix B). The structure of the $\boldsymbol{X}$ and $\boldsymbol{\omega}$ matrices is described and illustrated by *Hill and Tiedeman* [2007, p. 384]. In equation (11), the assumption has been made that the sample common error variance $\sigma^2 = 1$, hence, it does not explicitly appear in equation (11). This assumption is discussed toward the end of this appendix.

[73] For equation (11), as presented in this work, the terms for prior information are included in the matrices $\boldsymbol{X}$ and $\boldsymbol{\omega}$. Separation of $\boldsymbol{X}$ and $\boldsymbol{\omega}$ into the parts for observations and prior information produces the following alternative forms of equation (11):

$$\begin{aligned} V_L(\boldsymbol{\theta}) &= s^2(\boldsymbol{X}_{obs}^T\boldsymbol{\omega}_{obs}\boldsymbol{X}_{obs}+\boldsymbol{C}_p^{-1})^{-1} \\ &= \boldsymbol{C}_p-\boldsymbol{C}_p\boldsymbol{X}_{obs}(\boldsymbol{\omega}_{obs}^{-1}+\boldsymbol{X}_{obs}\boldsymbol{C}_p\boldsymbol{X}_{obs}^T)^{-1}\boldsymbol{X}_{obs}\boldsymbol{C}_p. \end{aligned} \tag{A1}$$

[74] The equivalence of the two expressions is discussed by *Oliver et al*. [2008, p. 142].

[75] The expression, $V(\boldsymbol{\theta})=\sigma^2(\boldsymbol{X}^T\boldsymbol{\omega}\boldsymbol{X})^{-1}$, is derived in detail by *Hill and Tiedeman* [2007, pp. 396–398], and a condensed version of that derivations follows.

[76] Start with the standard expression of variance applied to $\boldsymbol{\theta}$,

$$V(\boldsymbol{\theta})=E[(\boldsymbol{\theta}-E(\boldsymbol{\theta}))(\boldsymbol{\theta}-E(\boldsymbol{\theta}))^T]. \tag{A2}$$

[77] The expression for parameter values optimized using linear regression is

$$\boldsymbol{\theta}=(\boldsymbol{X}^T\boldsymbol{\omega}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\omega}\boldsymbol{y}, \tag{A3}$$

where $\boldsymbol{y}$ is a vector of observations. Substitution of equation (A3) into equation (A2) for $\boldsymbol{\theta}$ (not for E($\boldsymbol{\theta}$)); expansion of terms using matrix algebra; application of the matrix property $\boldsymbol{A}\boldsymbol{B}^T=\boldsymbol{B}^T\boldsymbol{A}^T$; recognition that $(\boldsymbol{X}^T\boldsymbol{\omega}\boldsymbol{X})^{-1}$ and $\boldsymbol{\omega}$ are symmetric so that the transpose equals the original matrix; substituting the first-order (linear) expression for $\boldsymbol{y}$, $\boldsymbol{y}=\boldsymbol{A}\boldsymbol{\theta}+\boldsymbol{\epsilon}$; and taking the expected value noting that only $\boldsymbol{\epsilon}$ is stochastic yields:

$$V_L(\boldsymbol{\theta})=(\boldsymbol{X}^T\boldsymbol{\omega}\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{\omega}E[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T]\boldsymbol{\omega}\boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{\omega}\boldsymbol{X})^{-1}. \tag{A4}$$

[78] The subscript L indicates linearity has been assumed. Consider defining the weight matrix ω such that

$$E[\boldsymbol{\epsilon}\boldsymbol{\epsilon}^T]=V(\boldsymbol{\epsilon})=\sigma^2\boldsymbol{\omega}^{-1}. \tag{A5}$$

[79] Substituting equation (A5) into equation (A4), canceling terms, and assuming $\sigma^2 = 1$ yields $V(\boldsymbol{\theta})=$

$(\boldsymbol{X}^T\boldsymbol{\omega}\boldsymbol{X})^{-1}$, the expression used in equation (A1). Assuming $\sigma^2 = 1$ is not significant for the DELSA statistics defined in this work because as a constant it divides out in the calculation of equation (13).

[80] The second part of equation (11) is the pre and post multiplication of $V_L(\boldsymbol{\theta})$ by the derivatives of the prediction with respect to the parameters. This multiplication is a linear propagation of the parameter error as represented by $V_L(\boldsymbol{\theta})$ to the prediction. Error propagation of this type is standard in regression theory, and is presented in the references cited for equation (11) at the beginning of this appendix.

[81] Equation (11) calculates the variances that appear in the denominator of equation (13) for the local first-order sensitivity indices. The numerator of equation (13) is obtained by retaining one term in the expression for $V_L(\boldsymbol{\theta})$, as described in the text.

Appendix B: Why the Variance of a Uniform Distribution Is Expected to Yield Numerically Equivalent DELSA and Sobol' Sensitivity Index Values Under Ideal Circumstances

[82] For the Sobol' sampling used in this work, the probability for all parameter values within the range of parameters is assumed to be equal. This is consistent with a uniform probability distribution.

[83] For the DELSA statistics (equation (13)) to produce numerically equivalent results to Sobol', the parameter range and distribution implied by the prior information variance in $\boldsymbol{\omega}$ needs to be consistent with that of the Sobol' method. Thus, given a parameter θ_j with ranges $\theta_{j,min}$ and $\theta_{j,max}$ defined for Sobol', the variance used in DELSA needs to equal the variance of the uniform distribution, which is $\frac{1}{12}(\theta_{j,max}-\theta_{j,min})^2$ [e.g., *Mood et al*., 1974].

[84] **Acknowledgments.** This study was done at NCAR while Olda Rakovec was visiting Martyn Clark. This work is financially supported by the Flood Control 2015 program (http://www.floodcontrol2015.nl), which is gratefully acknowledged. Mary C. Hill's involvement was supported by the USGS National Research Program, National Water Quality Assessment Program, and Groundwater Resources Program. We thank the Hydrological Service of the Walloon Region of Belgium (MET-SETHY) for providing the hydrological data, and we thank Marc Aubinet and Bernard Heinesch from The University of Liège for providing the meteorological data from the Vielsalm FLUXNET site. We also greatly appreciate the constructive comments from Hoshin V. Gupta of the University of Arizona, Tuscon; Emmanuele Borgonovo of Bocconi University, Milan; Sergei Kucherenko of Imperial College, London; Pablo Mendoza of NCAR, Boulder; two anonymous reviewers; and the Associate Editor, which certainly improved the quality of the manuscript.

References

Anderman, E. R., M. C. Hill, and E. P. Poeter (1996), Two-dimensional advective transport in ground-water flow parameter estimation, *Groundwater*, *34*(6), 1001–1009.

Archer, G., A. Saltelli, and I. Sobol (1997), Sensitivity measures, ANOVA-like techniques and the use of bootstrap, *J. Stat. Comput. Simul.*, *58*(2), 99–120, doi:10.1080/00949659708811825.

Aster, R. C., B. Borchers, and C. H. Thurber (2013), *Parameter Estimation and Inverse Problems*, Academic, Amsterdam.

Aubinet, M., B. Chermanne, M. Vandenhaute, B. Longdoz, M. Yernaux, and E. Laitat (2001), Long term carbon dioxide exchange above a mixed forest in the Belgian Ardennes, *Agric. For. Meteorol.*, *108*, 293–315.

Bedford, T. (1998), Sensitivity indices for (tree-) dependent variables, in *Proceedings of the Second International Symposium on Sensitivity Analysis of Model Output (SAMO98)*, edited by K. Chan, S. Tarantola, and F. Campolongo, pp. 17–20, EUR Report.

Borgonovo, E. (2006), Measuring uncertainty importance: Investigation and comparison of alternative approaches, *Risk Anal.*, *26*(5), 1349–1361, doi:10.1111/j.1539-6924.2006.00806.x.

Borgonovo, E. (2007), A new uncertainty importance measure, *Reliab. Eng. Syst. Safety*, *92*(6), 771–784, doi:10.1016/j.ress.2006.04.015.

Box, G. E. P., and G. C. Tiao (1992), *Bayesian Inference in Statistical Analysis*, Wiley, New York.

Bratley, P., and B. Fox (1988), Implementing Sobol's quasirandom sequence generator, *ACM Trans. Math. Software*, *14*(1), 88–100, doi:10.1145/42288.214372.

Chun, M.-H., S.-J. Han, and N.-I. Tak (2000), An uncertainty importance measure using a distance metric for the change in a cumulative distribution function, *Reliab. Eng. Syst. Safety*, *70*(3), 313–321, doi:10.1016/S0951-8320(00)00068-5.

Clapp, R. B., and G. M. Hornberger (1978), Empirical equations for some soil hydraulic properties, *Water Resour. Res.*, *14*(4), 601–604.

Clark, M., and D. Kavetski (2010), Ancient numerical daemons of conceptual hydrological modeling: 1. Fidelity and efficiency of time stepping schemes, *Water Resour. Res.*, *46*, W10510, doi:10.1029/2009WR008894.

Clark, M. P., A. G. Slater, D. E. Rupp, R. A. Woods, J. A. Vrugt, H. V. Gupta, T. Wagener, and L. E. Hay (2008), Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, *Water Resour. Res.*, *44*, W00B02, doi:10.1029/2007WR006735.

Clark, M. P., H. K. McMillan, D. B. G. Collins, D. Kavetski, and R. A. Woods (2011), Hydrological field data from a modeller's perspective: Part 2: Process-based evaluation of model hypotheses, *Hydrol. Processes*, *25*(4), 523–543, doi:10.1002/hyp.7902.

Cloke, H. L., F. Pappenberger, and J.-P. Renaud (2007), Multi-Method Global Sensitivity Analysis (MMGSA) for modelling floodplain hydrological processes, *Hydrol. Processes*, *22*(11), 1660–1674, doi:10.1002/hyp.6734.

Cook, R. D., and S. Weisberg (1982), *Residuals and Influence in Regression*, Chapman and Hall, New York.

Cukier, R. I., C. M. Fortuin, K. E. Shuler, A. G. Petschek, and J. H. Schaibly (1973), Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I. Theory, *J. Chem. Phys.*, *59*(8), 3873–3878, doi:10.1063/1.1680571.

Cukier, R. I., J. H. Schaibly, and K. E. Shuler (1975), Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. III. Analysis of the approximations, *J. Chem. Phys.*, *63*(3), 1140–1149, doi:10.1063/1.431440.

Cukier, R. I., H. B. Levine, and K. E. Shuler (1978), Nonlinear sensitivity analysis of multiparameter model systems, *J. Comput. Phys.*, *26*(1), 1–42, doi:10.1016/0021-9991(78)90097-9.

D'Agnese, F. A., C. C. Faunt, M. C. Hill, and A. Keith Turner (1999), Death valley regional ground-water flow model calibration using optimal parameter estimation methods and geoscientific information systems, *Adv. Water Resour.*, *22*(8), 777–790.

Delenne, C., B. Cappelaere, and V. Guinot (2012), Uncertainty analysis of river flooding and dam failure risks using local sensitivity computations, *Reliab. Eng. Syst. Safety*, *107*, 171–183, doi:10.1016/j.ress.2012.04.007.

Doherty, J., and D. Welter (2010), A short exploration of structural noise, *Water Resour. Res.*, *46*, W05525, doi:10.1029/2009WR008377.

Draper, N. R., and H. Smith (1998), *Applied Regression Analysis*, Wiley, New York.

Driessen, T. L. A., R. T. W. L. Hurkmans, W. Terink, P. Hazenberg, P. J. J. F. Torfs, and R. Uijlenhoet (2010), The hydrological response of the Ourthe catchment to climate change as modelled by the HBV model, *Hydrol. Earth Syst. Sci.*, *14*(4), 651–665, doi:10.5194/hess-14-651-2010.

Efron, B., and R. J. Tibshirani (1993), *An Introduction to the Bootstrap*, Chapman and Hall, New York.

Fienen, M. N., J. E. Doherty, R. J. Hunt, and H. W. Reeves (2010), Using prediction uncertainty analysis to design hydrologic monitoring networks: Example applications from the great lakes water availability pilot project, *U.S. Geol. Surv. Sci. Invest. Rep. 2010–5159*, 44 pp.

Foglia, L., S. W. Mehl, M. C. Hill, P. Perona, and P. Burlando (2007), Testing alternative ground water models using cross-validation and other methods, *Groundwater*, *45*(5), 627–641, doi:10.1111/j.1745–6584.2007.00341.x.

Foglia, L., M. C. Hill, S. W. Mehl, and P. Burlando (2009), Sensitivity analysis, calibration, and testing of a distributed hydrological model using error-based weighting and one objective function, *Water Resour. Res.*, *45*, W06427, doi:10.1029/2008WR007255.

Foglia, L., M. C. Hill, S. W. Mehl, and P. Burlando (2013), Evaluating model structure adequacy: The case of the Maggia Valley groundwater system, Southern Switzerland, *Water Resour. Res.*, *49*, 1–13, doi:10.1029/2011WR011779.

Freer, J., K. Beven, and B. Ambroise (1996), Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach, *Water Resour. Res.*, *32*(7), 2161–2173, doi:10.1029/95WR03723.

Göhler, M., J. Mai, and M. Cuntz (2013), Use of eigendecomposition in a parameter sensitivity analysis of the Community Land Model, *J. Geophys. Res. Biogeosci.*, *118*, 904–921, doi:10.1002/jgrg.20072.

Gupta, H. V., M. P. Clark, J. A. Vrugt, G. Abramowitz, and M. Ye (2012), Towards a comprehensive assessment of model structural adequacy, *Water Resour. Res.*, *48*, W08301, doi:10.1029/2011WR011044.

Hamm, N. A. S., J. W. Hall, and M. G. Anderson (2006), Variance-based sensitivity analysis of the probability of hydrologically induced slope instability, *Comput. Geosci.*, *32*(6), 803–817, doi:10.1016/j.cageo.2005.10.007.

Hazenberg, P., H. Leijnse, and R. Uijlenhoet (2011), Radar rainfall estimation of stratiform winter precipitation in the Belgian Ardennes, *Water Resour. Res.*, *47*, W02507, doi:10.1029/2010WR009068.

Helton, J. C. (1993), Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal, *Reliab. Eng. Syst. Safety*, *42*(2–3), 327–367, doi:10.1016/0951-8320(93)90097-I.

Herman, J. D., P. M. Reed, and T. Wagener (2013a), Time-varying sensitivity analysis clarifies the effects of watershed model formulation on model behavior, *Water Resour. Res.*, *49*, 1400–1414, doi:10.1002/wrcr.20124.

Herman, J. D., J. B. Kollat, P. M. Reed, and T. Wagener (2013b), Technical note: Method of Morris effectively reduces the computational demands of global sensitivity analysis for distributed watershed models, *Hydrol. Earth Syst. Sci.*, *17*, 2893–2903, doi:10.5194/hessd-10-4275-2013.

Hill, M. C., and C. R. Tiedeman (2007), *Effective Groundwater Model Calibration: With Analysis of Data, Sensitivities, Prediction and Uncertainty*, 455 pp., Wiley, Hoboken, N. J.

Homma, T., and A. Saltelli (1996), Importance measures in global sensitivity analysis of nonlinear models, *Reliab. Eng. Syst. Safety*, *52*(1), 1–17, doi:10.1016/0951-8320(96)00002–6.

Hornberger, G., and R. Spear (1981), An approach to the preliminary analysis of environmental systems, *J. Environ. Manage.*, *12*(1), 7–18.

Katz, R. W. (1999), Extreme value theory for precipitation: Sensitivity analysis for climate change, *Adv. Water Resour.*, *23*(2), 133–139, doi:10.1016/S0309-1708(99)00017-2.

Kavetski, D., and M. Clark (2010), Ancient numerical daemons of conceptual hydrological modeling: 2. Impact of time stepping schemes on model analysis and prediction, *Water Resour. Res.*, *46*, W10510, doi:10.1029/2009WR008896.

Kavetski, D., and G. Kuczera (2007), Model smoothing strategies to remove microscale discontinuities and spurious secondary optima in objective functions in hydrological calibration, *Water Resour. Res.*, *43*, W03411, doi:10.1029/2006WR005195.

Kleijnen, J. P. C., and J. C. Helton (1999), Statistical analysis of scatterplots to identify important factors in large-scale simulations, 1: Review and comparison techniquest, *Reliab. Eng. Syst. Safety*, *65*(2), 147–185.

Kucherenko, S., M. Rodriguez-Fernandez, C. Pantelides, and N. Shah (2009), Monte Carlo evaluation of derivative-based global sensitivity measures, *Reliab. Eng. Syst. Safety*, *94*(7), 1135–1148, doi:10.1016/j.ress.2008.05.006.

Kucherenko, S., S. Tarantola, and P. Annoni (2012), Estimation of global sensitivity indices for models with dependent variables, *Comput. Phys. Commun.*, *183*(4), 937–946, doi:10.1016/j.cpc.2011.12.020.

Kunstmann, H., W. Kinzelbach, and T. Siegfried (2002), Conditional first-order second-moment method and its application to the quantification of uncertainty in groundwater modeling, *Water Resour. Res.*, *38*(4), doi:10.1029/2000WR000022.

Li, J., Q. Y. Duan, W. Gong, A. Ye, Y. Dai, C. Miao, Z. Di, C. Tong, and Y. Sun (2013), Assessing parameter importance of the Common Land Model based on qualitative and quantitative sensitivity analysis, *Hydrol. Earth Syst. Sci.*, *17*(8), 3279–3293, doi:10.5194/hess-17-3279-2013.

Lu, D., M. Ye, and M. C. Hill (2012), Analysis of regression confidence intervals and Bayesian credible intervals for uncertainty quantification, *Water Resour. Res.*, *48*, W09521, doi:10.1029/2011WR011289.

Massmann, C., and H. Holzmann (2012), Analysis of the behavior of a rainfall-runoff model using three global sensitivity analysis methods evaluated at different temporal scales, *J. Hydrol.*, *475*, 97–110, doi:10.1016/j.jhydrol.2012.09.026.

Mehl, S. W., and M. C. Hill (2001), A comparison of solute-transport solution techniques and their effect on sensitivity analysis and inverse modeling results, *Groundwater*, *39*(2), 300–307.

Menke, W. (1989), *Geophysical Data Analysis, Discrete Inverse Theory, Int. Geophys. Ser.*, vol. 45, Academic, New York.

Mood, A. M., F. A. Graybill, and D. C. Boes (1974), *Introduction to the Theory of Statistics*, 3rd ed., McGraw-Hill, New York.

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

Musters, P. A. D., and W. Bouten (2000), A method for identifying optimum strategies of measuring soil water contents for calibrating a root water uptake model, *J. Hydrol.*, *227*, 273–286.

Nossent, J., P. Elsen, and W. Bauwens (2011), Sobol' sensitivity analysis of a complex environmental model, *Environ. Modell. Software*, *26*(12), 1515–1525, doi:10.1016/j.envsoft.2011.08.010.

Oakley, J. E., and A. O'Hagan (2004), Probabilistic sensitivity analysis of complex models: A Bayesian approach, *J. R. Stat. Soc., Ser. B*, *66*(3), 751–769, doi:10.1111/j.1467-9868.2004.05304.x.

Oliver, D. S., A. C. Reynolds, and N. Liu (2008), *Inverse Theory for Petroleum Reservoir Characterization and History Matching*, Cambridge Univ. Press, New York, N. Y.

Papale, D., et al. (2006), Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: Algorithms and uncertainty estimation, *Biogeosciences*, *3*(4), 571–583, doi:10.5194/bg-3–571-2006.

Pappenberger, F., P. Matgen, K. J. Beven, J. B. Henry, L. Pfister, and P. Fraipont (2006), Influence of uncertain boundary conditions and model structure on flood inundation predictions, *Adv. Water Resour.*, *29*(10), 1430–1449.

Pappenberger, F., K. J. Beven, M. Ratto, and P. Matgen (2008), Multimethod global sensitivity analysis of flood inundation models, *Adv. Water Resour.*, *31*(1), 1–14, doi:10.1016/j.advwatres.2007.04.009.

Park, C. K., and K.-I. Ahn (1994), A new approach for measuring uncertainty importance and distributional sensitivity in probabilistic safety assessment, *Reliab. Eng. Syst. Safety*, *46*(3), 253–261, doi:10.1016/0951-8320(94)90119-8.

Plischke, E., E. Borgonovo, and C. L. Smith (2013), Global sensitivity measures from given data, *Eur. J. Oper. Res.*, *226*(3), 536–550, doi:10.1016/j.ejor.2012.11.047.

Rakovec, O., A. H. Weerts, P. Hazenberg, P. J. J. F. Torfs, and R. Uijlenhoet (2012), State updating of a distributed hydrological model with Ensemble Kalman Filtering: Effects of updating frequency and observation network density on forecast accuracy, *Hydrol. Earth Syst. Sci.*, *16*, 3435–3449, doi:10.5194/hess-16-3435-2012.

Reichstein, M., et al. (2005), On the separation of net ecosystem exchange into assimilation and ecosystem respiration: Review and improved algorithm, *Global Change Biol.*, *11*(9), 1424–1439, doi:10.1111/j.1365-2486.2005.001002.x.

Rosolem, R., H. V. Gupta, W. J. Shuttleworth, X. Zeng, and L. G. G. De Gonçalves (2012), A fully multiple-criteria implementation of the Sobol' method for parameter sensitivity analysis, *J. Geophys. Res. Atmos.*, *117*, D07103, doi:10.1029/2011JD016355.

Saltelli, A. (2002), Making best use of model evaluations to compute sensitivity indices, *Comput. Phys. Commun.*, *145*(2), 280–297, doi:10.1016/S0010-4655(02)00280-1.

Saltelli, A., S. Tarantola, and K. P.-S. Chan (1999), A quantitative model-independent method for global sensitivity analysis of model output, *Technometrics*, *41*(1), 39–56.

Saltelli, A., S. Tarantola, F. Campolongo, and M. Ratto (2004), *Sensitivity Analysis in Practice*, Wiley, New York.

Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola (2008), *Global Sensitivity Analysis: The Primer*, Wiley, West Sussex, U. K.

Seber, G. A. F., and C. J. Wild (1989), *Nonlinear Regression*, Wiley, New York.

Sobol', I. M. (2001), Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Math. Comput. Simul.*, *55*(1–3), 271–280, doi:10.1016/S0378-4754(00)00270-6.

Sobol', I. M., and S. Kucherenko (2009), Derivative based global sensitivity measures and their link with global sensitivity indices, *Math. Comput. Simul.*, *79*(10), 3009–3017, doi:10.1016/j.matcom.2009.01.023.

Sobol', I. M., and S. Kucherenko (2010), A new derivative based importance criterion for groups of variables and its link with the global sensitivity indices, *Comput. Phys. Commun.*, *181*, 1212–1217.

Tang, Y., P. Reed, T. Wagener, and K. van Werkhoven (2007), Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation, *Hydrol. Earth Syst. Sci.*, *11*(2), 793–817, doi:10.5194/hess-11-793-2007.

Tarantola, A. (2005), *Inverse Problem Theory and Methods for Model Parameter Estimation*, Soc. for Ind. and Appl. Math., Philadelphia, Pa.

Tiedeman, C. R., M. C. Hill, F. A. D'Agnese, and C. C. Faunt (2003), Methods for using groundwater model predictions to guide hydrogeologic data collection, with application to the Death Valley regional groundwater flow system, *Water Resour. Res.*, *39*(1), 1010, doi:10.1029/2001WR001255.

Tiedeman, C. R., D. M. Ely, M. C. Hill, and G. M. O'Brien (2004), A method for evaluating the importance of system state observations to model predictions, with application to the Death Valley regional groundwater flow system, *Water Resour. Res.*, *40*, W12411, doi:10.1029/2004WR003313.

Tonkin, M. J., C. R. Tiedeman, D. M. Ely, and M. C. Hill (2007), OPR-PPR, A computer program for assessing data importance to model predictions using linear statistics, *U.S. Geol. Surv. Tech. and Methods Rep. TM-6E2*, 115 pp.

van Griensven, A., T. Meixner, S. Grunwald, T. Bishop, M. Diluzio, and R. Srinivasan (2006), A global sensitivity analysis tool for the parameters of multi-variable catchment models, *J. Hydrol.*, *324*(1–4), 10–23, doi:10.1016/j.jhydrol.2005.09.008.

Van Werkhoven, K., T. Wagener, P. Reed, and Y. Tang (2008), Characterization of watershed model behavior across a hydroclimatic gradient, *Water Resour. Res.*, *44*, W01429, doi:10.1029/2007WR006271.

Vrugt, J., W. Bouten, and A. Weerts (2001), Information content of data for identifying soil hydraulic parameters from outflow experiments, *Soil Sci. Soc. Am. J.*, *65*(1), 19–27.

Wagener, T., N. McIntyre, M. J. Lees, H. S. Wheater, and H. V. Gupta (2003), Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis, *Hydrol. Processes*, *17*(2), 455–476, doi:10.1002/hyp.1135.

Weerts, A., J. Huisman, and W. Bouten (2001), Information content of time domain reflectometry waveforms, *Water Resour. Res.*, *37*(5), 1291–1299, doi:10.1029/2000WR900381.

Xu, C., and G. Z. Gertner (2011), Reliability of global sensitivity indices, *J. Stat. Comput. Simul.*, *81*(12), 1939–1969, doi:10.1080/00949655.2010.509317.

Ye, M., K. F. Pohlmann, J. B. Chapman, G. M. Pohll, and D. M. Reeves (2010), A model-averaging method for assessing groundwater conceptual model uncertainty, *Groundwater*, *48*(5), 716–728, doi:10.1111/j.1745–6584.2009.00633.x.

Zambrano-Bigiarini, M., and R. Rojas (2013), A model-independent Particle Swarm Optimisation software for model calibration, *Environ. Modell. Software*, *43*, 5–25, doi:10.1016/j.envsoft.2013.01.004.

Zhan, C.-S., X.-M. Song, J. Xia, and C. Tong (2013), An efficient integrated approach for global sensitivity analysis of hydrological model parameters, *Environ. Modell. Software*, *41*, 39–52, doi:10.1016/j.envsoft.2012.10.009.

Zuniga, M. M., S. Kucherenko, and N. Shah (2013), Metamodelling with independent and dependent inputs, *Comput. Phys. Commun.*, *184*(6), 1570–1580, doi:10.1016/j.cpc.2013.02.005.